12-13-2003

An Investigation of the Standardized Multiple-Choice Departmental Calculus I Final Examination

Maria Elizabeth Bearden

AN INVESTIGATION OF THE STANDARDIZED MULTIPLE-CHOICE DEPARTMENTAL CALCULUS I FINAL EXAMINATION

By

Maria Elizabeth Bearden

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Education
in the Department of Curriculum and Instruction

Mississippi State, Mississippi

December 2003

AN INVESTIGATION OF THE STANDARDIZED MULTIPLE-CHOICE

DEPARTMENTAL CALCULUS I FINAL EXAMINATION

By

Maria Elizabeth Bearden

Approved:

R. Dwight Hare
Professor of Curriculum and Instruction
(Director of Dissertation)

Jerry F. Reed
Professor Emeritus of Mathematics
and Statistics
(Minor Professor)

David T. Morse
Professor of Counselor Education
and Educational Psychology
(Committee Member)

Sue S. Minchew
Assistant Dean of The College of Education
Professor of Curriculum and Instruction
(Committee Member)

Dana Lea Franz
Assistant Professor of Curriculum
and Instruction
(Committee Member)

Linda T. Coats
Associate Professor of Curriculum
and Instruction
Graduate Coordinator

Roy H. Ruby
Dean of the College of Education

Name: Maria Elizabeth Bearden

Date of Degree: December 13, 2003

Institution: Mississippi State University

Major Field: Secondary Education (Mathematics)

Major Professor: Dr. R. Dwight Hare

Title of Study: AN INVESTIGATION OF THE STANDARDIZED MULTIPLE-CHOICE DEPARTMENTAL CALCULUS I FINAL EXAMINATION

Pages in Study: 150

Candidate for the Degree of Doctor of Philosopy

At Mississippi State University, Department of Mathematics and Statistics, a standardized multiple-choice departmental final examination (SMCDF) was administered at the end of the Calculus I mathematics course. This practice was abolished at the end of the spring 1997 semester. The purpose of this study was to determine if there was a difference in students' success in subsequent calculus courses as measured by a student's grade. If there was a difference, was it consistent along varying levels of students' ACT mathematics scores.

A two-way analysis of variance (ANOVA) was run on the data. The variables were the five student ACT Mathematics Standards for Transition ranges, and the two groups of students, those required to take a SMCDF examination and those required to take a teacher generated final examination. The results showed there was a significant difference in the mean grades at the Calculus II level (p=.006), suggesting the SMCDF

examination in Calculus I improved their level of success in Calculus II. In the Calculus III and IV courses no significant differences were found.

When descriptive statistics were analyzed, an unusual number of F grades were found in one group due to a university audit policy that was abolished in the fall of 1997. When F grades were excluded from the data, no significant differences were found for Calculus II, III, or IV. Further investigation along ACT-Mathematics Standards for Transition ranges showed, at an alpha level of .01 for Calculus II and IV, the data set was too small at each of these ranges to determine any significant differences.

Although conflicting results did not clearly indicate whether a SMCDF examination made a difference, indications seem to be at least at the Calculus II level there was a significant effect in the original data set. Descriptive statistics showed inconsistencies within the Calculus III data as compared to Calculus II and IV.

Further investigation was recommended for this area of research. Incorporating teaching styles into this study and changing the format of the examination were suggested.

DEDICATION

I would like to dedicate this research to Dr. Jerry Reed who encouraged me to pursue this goal, had faith in me when I questioned myself, and for always willing to give me such sound advice. Thanks for being my friend.

ACKNOWLEDGMENTS

I would like to thank all the members of my graduate committee: Dr. Sue S. Minchew, Dr. Dana Lea Franz, Dr. David T. Morse, Dr. Jerry F. Reed, and especially Dr. Dwight Hare, my major professor and dissertation director. Thank you for all your support. Dr. Minchew, thank you for listening to my frustrations, reviewing my work, and giving me words of encouragement. Dr. Franz, even though you came aboard at a later stage, I appreciate your willingness to jump in and help, and give me suggestions to improve on my project. Dr. Morse, I thank you for agreeing to take on one more graduate student by serving on my committee and advising me with the statistics. Your class taught me so much about psychological statistics; it all became much more meaningful when I had to apply it in my dissertation. I appreciate all your help. Dr. Reed, even though I have never taken a class under you, I have learned many lessons from you just by your example as a teacher and a professional while I worked with you in the mathematics department. I appreciate you always being there and helping me when needed. Your words of encouragement, your recommendations and your belief in me have carried me to complete this degree. I must also thank Dr. Pat Gerard and Dr. Regina Halpin who served on my committee during written and oral exams, and the early stages of the dissertation. Dr. Gerard, thank you for your help with the statistics even though your background was not in statistics for educational research. You gave me support and statistical advice when I needed it. Dr. Halpin, you instructed me in educational research

and started out as my major professor. I appreciate your guidance and willingness to continue advising me even after you and your family moved away. Thank you for always encouraging me and for all the work you have done. A special thank you goes to Mr. John T. O'Bannon from ITS Information System and his staff. Without his programming skills and persistence, I would have never been able to get the data to do this research. Finally, Dr. Hare, thank you so much for taking over when Dr. Halpin moved away. You were already advising many other graduate students, teaching classes, conducting your own research, and otherwise overworked, yet you were willing to take me on board. I would never have been able to complete this dissertation without your help. Looking back on the process, I must say that you have made me learn a lot about how to conduct a research project. I can now see how your method of answering a question with a question, as frustrating as it can be, has made me learn to ask these questions of myself. I must thank you for the many drafts that you have read and marked in so many places, which in turn have made me proud of my results. I must confess that once there was a time when I did not like the research process, but this experience, as tedious and time consuming as it can be, has made me appreciate the work and actually look forward to being able to work on other projects. I thank you for that.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Ferrini-Mundy and Graham (1991), over 10 years ago, noted that about 600,000 students would enroll in a beginning level calculus course in four-year colleges and universities throughout the United States each year. A large number of these students were required to study calculus as an integral part of their major area of study in the anticipation of pursuing their chosen occupational field. Specifically, about half of these students were enrolled in mainstream engineering calculus. Over the years, calculus has continued to play an important role in engineering, business and other disciplines of study. According to Wilson (1997), mathematics departments often depend on and bring in more tuition dollars for students that are in calculus courses than any other mathematics course.

Calculus for engineering or physics-based calculus is generally taught as a three or four semester series of courses with the first semester being a prerequisite for the subsequent semesters (Bulletin of Mississippi State University, 1997). The beginning level of calculus, often referred to as Calculus I or Differential Calculus, is the foundation on which the following courses rely. A poor performance at this level more than likely will affect the level of performance in subsequent courses. Furthermore, this may also affect decisions in continuing the intended field of study (Andrade, 1999; Millar and others, 1996). According to a study by Bashford (2000), passing a prerequisite course does not necessarily prepare the student for the next course in the sequence. Success at the beginning level of calculus as determined by a grade or promotion to the next level does not necessarily mean the student is prepared for the next level of calculus.

Since there are so many students registering for the beginning level of calculus, institutions of higher learning will offer multiple sections of calculus each semester. Professors, instructors, or graduate students may teach these classes; thus, experience in teaching may vary significantly amongst these individuals. When a number of different teachers are available to teach the same material, an issue of consistency in the depth of understanding the material covered will likely arise (Seese, 1994).

Purpose of the Study

Prior to the fall of 1997, the Mississippi State University (MSU) Department of Mathematics and Statistics required a standardized multiple-choice departmental final (SMCDF) examination at the end of the Calculus I mathematics course. This examination, which was implemented in 1983, was to appraise the competence of all students over a predetermined set of skills. At the end of the spring 1997 semester, this practice was discontinued. Commencing with the fall of 1997, each teacher in the beginning Calculus I course was still informed of the material that should be covered, but the rigor of the course and the nature of the final examination were left to each teacher's discretion. Thus, the content and format of the final examination, and thereby, the competency level to be demonstrated by the student, were left solely to the teacher. The purpose of this study was to determine if there is a difference in students' success in subsequent calculus courses before and after use of the standardized multiple-choice departmental final (SMCDF) examination for Calculus I.

Research Question

Is there a difference in student's success in subsequent calculus courses as measured by a student's grade in the course, when there is a required standardized multiple-choice departmental final (SMCDF) examination at the end of the Calculus I course as compared to a required final examination generated by the individual instructor at the end of Calculus I? If there is a difference, is this difference consistent along varying levels of students' ACT mathematics assessment scores?

Rationale of the Study

Many of our engineering and science fields require students to have a solid background in mathematics. This usually includes the required four-semester sequence of courses generally referred to as Calculus I, Calculus II, Calculus III, and Calculus IV. The first calculus course becomes the foundation for successive courses. When different teachers are assigned the task of preparing these students, the level of competency at the end of the term may vary considerably. In order to bring some uniform level of success in all sections being taught, teachers assigned to instruct the beginning calculus course should make sure that all students have been given the opportunity to learn the course objectives and can demonstrate they are properly prepared to further their learning in subsequent calculus courses. Traditionally, when university students select their classes, they inquire about the teachers listed in the class schedule before making their selection. The decision to select a particular section may be based upon the amount of work expected in the course or the grading policies of the teacher. The most important question not asked is whether the class will prepare them well for subsequent calculus courses.

Students may have a preference in teaching and learning styles, but the level of competence after the first course in calculus must be such that students can be successful in subsequent calculus courses. This study focused on only one aspect of a specific course, the final examination. The researcher investigated the effect of changing a uniform final examination policy for Calculus I.

Definitions of Terms

The following definitions will be used in this study:

1. ACT mathematics assessment score – score from the mathematics portion of the American College Testing program designed to assess high school students' general educational development and their ability to complete college-level work.

2. ACT's Standards for Transition - sets of statements that describe the academic knowledge and skills those students have likely acquired. Developed by the American College Testing Program, standards are provided for six ACT score ranges (13-15, 16-19, 20-23, 24-27, 28-32, 33-36).

3. Beginning level Calculus/Calculus I/Differential Calculus/MA-1713 - first course in the calculus sequence. At MSU this consists of a 3 credit hour course.

4. <u>Standardized multiple-choice departmental final (SMCDF) examination</u> - an examination that is administered to all students simultaneously at the end of the course. In this study it refers to a 40 question multiple-choice examination that was administered to all Mississippi State University Calculus I students at the end of that course. The examination was written by a committee of professors from the MSU Department of Mathematics and Statistics. It was abolished in the fall semester of 1997.

5. <u>Teacher-generated final examination</u> – an examination that is administered to a single class or section at the end of the course. The examination is constructed by the person teaching that particular class.

6. <u>Norm referenced test</u> – a test that provides information about how a student's performance compares with a reference group instead of how many questions were answered correctly (Cizek, 1998; Haney & Madaus, 1989; Madaus, & O'Dwyer, 1999). Examples of such tests are the Otis-Lenon School Aptitude Test, Stanford-Binet, SAT, American College Testing (ACT) and the Iowa Test of Basic Skills (ITBS).

7. <u>Criterion referenced test</u> – a test that intends to measure how well a person has learned a specific body of knowledge and skills. Tests developed by the classroom teachers and student competency tests developed and used by individual states for purposes of determining grade-to-grade promotion or graduation are examples of criterion referenced tests.

8. <u>Formative assessment</u> – when testing is used for obtaining information for decision-making in the area of placement, diagnostics, attainment or monitoring progress.

When applied to students, formative assessment occurs at the classroom level for purpose of improving teaching and learning and focuses on particular assignments and concepts. The teacher has control over the questions. Formative assessment can be administered throughout the semester.

9. Summative assessment – when testing is used for determining how well a program worked, to provide a macro view of teaching, learning, and institutional effectiveness. The goal is to evaluate faculty for personnel decisions and focuses on the entire course. The institution or department has control over the questions. Summative assessment is administered at the end of the semester.

CHAPTER II

REVIEW OF THE LITERATURE

Quiz, test, and exam(ination) are all words that are quite familiar to students in any classroom. In educational literature, these words come under the title of assessment. Final examinations are common instruments of assessment for many college courses. This chapter will begin with defining assessment in basic terms. Purposes for assessment in education are discussed and instruments used in assessment are also described. Since the standardized multiple-choice departmental final (SMCDF) examination in Calculus I is the focus of this study, some background information on multiple-choice testing is presented next. This history gives a rationale for not only the purposes for varying types of assessment but also the different types of instruments used in the assessment process. The problems and limitations of assessment, from the instrument used in assessment to teacher training, are documented in the third section.

Calculus I as a prerequisite to the next course in the sequence is often referred to as a gatekeeper course. The fourth section of this review describes the beginning level calculus course as a gatekeeper course for science, mathematics, engineering and technology (SMET) majors. Since ACT mathematics scores are also used in this study, the fifth section gives an overview of the ACT Assessment program to help understand the possible significant role it may have in this study. Finally, since no studies were found that investigated the link between the SMCDF examination in a calculus class and success in subsequent courses in the calculus sequence, research that is somewhat related to this study is reported in the final section of this chapter.

Assessment Defined

In educational literature, the words quiz, test, examination all come under the realm of assessment. Steen (1999) expressed that assessment is an act of making an official evaluation of students' performance for the purpose of measuring what they know. In mathematics education, the National Council of Teachers of Mathematics (NCTM) (2000) described the assessment principle in its publication The Principles and Standards for School Mathematics. The principle simply stated that assessment should support the learning of important mathematics and furnish useful information to both teachers and students. William and Black (1996) explained that the information obtained is used for decision-making in the area of placement, diagnostics, attainment or monitoring progress and is often referred to as formative assessment. Formative assessment often focuses on particular assignments and lectures. This type of assessment can be administered throughout the semester or at the end of the course, the instructor has control over the questions. Over the years, however, assessment has been used for purposes that extend beyond this realm. In an article by Steen, assessment was also used to judge teaching effectiveness, to determine raises and promotions, to evaluate curricula and programs and to decide on allocation of resources. This type of assessment is referred to as summative assessment and is often controlled by an institution or department. Summative assessment has a broad focus on the entire course or program and is administered at the end of a semester or particular sequence of activities. Summative assessment enables the institution to make more effective use of resources and to enhance its accountability to internal and external constituencies.

Educational assessment was not just important for students, faculty and administrators, but was also found valuable and informative by parents, legislators, journalists and the public. Just as there were many reasons for assessment to occur, there were also many different testing instruments to accomplish this task. These instruments were in the form of series of questions, exercises, performances or other means of measuring the skill, knowledge, intelligence, capacities, or aptitude of an individual or group. Romagnano and Long (2001) wrote that the technique or format of the assessment instrument might have been of writing samples, oral questions, multiple-choice questions, open-ended problems, performance tasks, writing assignments, or portfolios.

Multiple Choice Testing

At the end of the first decade of the 20^{th} century, written essay tests, which had been the principal mode of examinations during the previous century, came under severe challenge. This was partly a response to studies that showed that the marks assigned to essay questions were highly unreliable. It was also a response to the school-based efficiency movement which required that growing numbers of students be tested to measure a school district's competence. With an increase in the number of examinees also came an increase in the labor for administering the examinations and recording the results (Haney & Madaus, 1989). This evolved in the need for more efficient, more manageable, and more easily recorded methods of organization. The use of essay examinations began to decline, and was slowly replaced by the short-answer format of testing.

In 1903, Frederick Kelly (as cited in Madaus & O'Dwyer, 1999) influenced researchers to come up with educational standards by which students could easily and scientifically be classed for informative purposes. This led to what we refer to as a norm-referenced test. Later in 1914 Kelly introduced the multiple-choice item. This format greatly facilitated the development of the national, norm-referenced, and standardized commercial tests. The Stanford-Binet, developed in 1910, was the first widely administered method of gauging human intelligence. In the 1930s a national association of educational institutions and organizations called the College Board implemented a college admissions test called the Scholastic Aptitude Test, also known as the SAT. It was designed to help colleges evaluate an applicant's chances of success in the freshman year, but quickly turned into an important national measure of school accountability (Linn, 2001; Stiggins, 1999a). Many students will retake the SAT several times over to improve their scores, driven by the idea that a high SAT score is associated with success and intelligence, which is often encouraged and reinforced by the media. A recent newspaper cover story article about the SAT referred to this exam as the one that Americans love to hate (Marklein, 2002).

In 1955 Everet Lindquist (as cited in Madaus & O'Dwyer, 1999) invented the high-speed optical scanner, which made multiple-choice tests even more efficient and cost effective. For the next 35 years this format of testing took precedence over any others since it could be scored in a fraction of the time, cost, better objectivity and increased reliability of scoring. In 1959 Lindquist also established the American College Testing (ACT) Assessment Program. He believed that the test should measure what students can do with what they have learned throughout their schooling. Many colleges and universities have used both the ACT and the SAT in making admission decisions, providing academic advisement, and career planning. These tests have also been used in the areas of recruitment and retention (Morgan, 1992).

The term “high stakes testing” is used when important sanctions or rewards are attached to test results (Kohn, 2000). Decisions about promotion or retention, placement in programs, certification of completion, teacher credentialing, allocation of funds or resources, merit pay, etc. are often based on the results of these standardized tests. Changes in assessment technology over the last two centuries have led to the streamlined, machine-scoreable, standardized, multiple-choice test as the technology of choice for policy makers in the United States during the 1970’s and 1980’s.

In the mid-1980s, many practitioners began to see the counter-productivity of this type of testing (Kohn, 2000). Madaus and Kellaghan (1993) referred to a number of sources, including the American Educational Research Association, that recognized that the testing programs of the 1970s and 1980s were having a negative effect on students, teachers, instruction, and learning. Their main argument was that multiple-choice, norm-referenced standardized tests used for policy purposes frequently measured low-level knowledge and skills, thus driving the system in the wrong direction, corrupting instruction and eventually the tests themselves. The ability to read fluently, write, speak, and reason began to suffer as educators under pressure prepared students for these tests. A number of researchers found that accountability pressures encouraged teachers and administrators to focus planning and instructional effort on test content. Teachers would translate all of their achievement targets into multiple-choice test items for assessment. Any concepts that could not be translated into multiple-choice test items would not be worth teaching because they could not be dependably and scientifically assessed (Haney & Madaus, 1989; Kohn, 2000; Stiggins, 1995). The standardized multiple-choice test came under criticism.

In the late 1980s, the education community began to focus on alternatives to the multiple-choice format of testing. Alternative assessment required the examinees to construct their own responses. This led to a resurgence of interest in performance assessment where the individual had to perform, create, or produce something over a sufficient amount of time to allow evaluation of either the process or the product, or both (Messick, 1994). The significance of the change to this alternative format was to focus on higher-order thinking and problem-solving skills. The era of assessment reform thus began. Performance assessment became the trendy phrase of the 1990s (Stiggins, 1995). Airasian (1997), however, pointed out that alternative testing is not really new to the education scene. Prior to the introduction of multiple-choice test items, educational assessments required that students construct their own responses. What had changed was the use of common, large-scale alternative assessments on a national, state- or district-wide basis, standardized scoring schemes and the consequential rewards or penalties for students and teachers linked to student performance on the assessments.

Many changes in assessment practices were evident in the classrooms. These changes included more frequent use of open-ended questions, portfolios, and performance assessment. Cizek (1993b) noted that Haney and Madaus of the Center of the Study of Testing, Evaluation, and Educational Policy at Boston College pointed out that performance assessment and other alternative evaluation methods that had been suggested over the last 20 years were not innovative. Evaluation tools such as live performances, products, teacher judgment, and school grades have had a long history of use in education.

Problems in Assessment

Assessment methods and instruments have strengths and weaknesses. The ultimate objective is to find the best assessment technique or instrument possible for the intended purpose. The purposes of assessment include validating learning progress, acceptance or placement into a program, certification, funding acquisition, ranking, and accountability. Problems may occur in the method of assessment, the instrument, in the evaluator, or in the purpose of the assessment itself. Some of these areas are examined below.

For decades the multiple-choice format of testing has been a dominant tool in the area of assessment (Cizek, 1998). This format of assessment, along with matching and true/false questions, is classified as selected-response test questions. They are used especially when assessing large groups of students in differing locations. A standardized multiple-choice examination provides consistency, is very time efficient, is easily administered, and is the most cost-effective format of assessment possible. However, there are also drawbacks. This type of test cannot efficiently evaluate reading fluency, writing ability, speaking, and reasoning skills. Madaus and Kellaghan (1993) and Herman (1992) listed a number of research references for the argument that the multiple-choice, norm-referenced standardized tests used for policy purposes often measure low-level knowledge and thinking skills. Kohn (2000) reported a study published in the Journal of Educational Psychology indicating that standardized-test results such as the SAT, ACT, ITBS, were positively correlated with a shallow approach to learning. As a result, education is being driven in the wrong direction. Furthermore, teachers feeling the pressure of accountability often will teach to the test (Haney & Madaus, 1989; Madaus & Kellaghan, 1993; Madaus & O'Dwyer, 1999). Cizek (1998) credited W. James Popham for the term <u>measurement driven instruction</u> (MDI) that is often used in reference to the power of mandated tests that influences teaching. Airasian (1997) also points out that teachers prepping students for tests and a narrowing curriculum may have more to do with the importance of the test consequences for students and teachers than the format of the assessment instrument.

In contrast to the multiple-choice (selected response) format of testing, there is the constructed-response test format. The essay, short-answer, speech, project, portfolio, journal, etc. are considered constructed-response types of tests (Cizek, 1998). Performance assessment requires examinees to construct/supply oral or written answers, perform or produce something for evaluation. It does not ask the examinee to select an answer to a question as in the multiple-choice format. Performance assessment is often viewed as a new approach, but as stated earlier, it was one of the earliest forms of assessment used. What is new is how performance assessment is now used to evaluate higher-order thinking skills. Stiggins (1995) preferred to describe these skills as the bonded relationship that exists between mastery of knowledge and mastery of reasoning. The knowledge and skills being evaluated involve student constructions, products, problem solving, and explanations (Peressini & Bassett, 1996). Small-scale alternative or performance assessment can be an instructional tool for a classroom teacher, but for large groups of examinees large-scale alternative assessment is less efficient, more difficult to administer, more disruptive to the school organization, and more time-consuming than multiple-choice testing programs (Baxter, Shavelson, Herman, Brown & Valadez, 1993; Cooney, Bell, Fisher-Cauble & Sanchez, 1996; Madaus & O'Dwyer, 1999; Miller & Linn, 2000). Assessing high- or low-level thinking skills may be a difficult task in itself (Arasian, 1997; Madaus & O'Dwyer, 1999). A high score in a performance task may be evidence that the examinee has mastered the content knowledge, reasoned productively, and performed skillfully. A low score may not indicate whether it is content, reasoning or performance that is lacking; therefore, a follow up with a different method of assessment might be helpful (Stiggins, 1995).

Another limitation of a performance-based test is the degree of objectivity that can be attained in such assessment method. An individual or group panel may be the evaluator. To be objective, the evaluator must be completely uninfluenced by emotion or personal prejudice (Romagnano & Long, 2001). Educational assessment is a process by which humans try to determine what other humans know; total objectivity is impossible. All assessments of students' understanding are subjective to some degree, as they can only be observed indirectly. A certain amount of objectivity is attained if predetermined criteria are set for the panel to use in their evaluation. An example of such a test is the free response section of the College Board Advanced Placement (AP) test that is widely used in the Advanced Placement Program. The AP panel of evaluators is trained for consistency in grading in an attempt to be more objective (Romagnano & Long, 2001). This, however, also contributes to the financial costs (Steen, 1999) and time requirements (Stiggins, 1995) of this type of assessment.

Teacher training can be viewed as a limitation in the assessment area of education. To be a certified teacher one must follow an approved curriculum of studies.

Assessment may or may not be a topic covered in the curriculum (Cizek, 1993b). Demonstrating competence in assessment as a condition to being licensed to teach is rarely a requirement. Hiring standards or staff evaluations are seldom based on the assessment capabilities of the teacher (Stiggins, 1995, 1999b). As a result, assessment or testing by the classroom teacher is very inconsistent from teacher to teacher and school to school. This is especially true for the less experienced teachers. A further limitation is that assessment training programs or in-school resources are very limited or non-existent in many areas. English researchers Black and William (as cited in Stiggins,1999a) reported consistent and sizable gains in elementary and high school students' standardized test scores directly attributed to improved teacher classroom assessment practices.

Kohn (2000) wrote about some problems in assessment. In particular, he stated norm-referenced tests may be biased because some questions require a set of knowledge and skills more likely to be possessed by students from a privileged background. He also pointed out that the more affluent schools and families can afford test-preparation materials and services to help in preparation. Examples of such tests are so-called intelligence, or I.Q., tests. These tests were thought of as a viable tool for classifying people in terms of their intelligence in the 1920s. David Lohman, in an article by Peter Airasian (1997) stated that intelligence tests are often viewed as biased and unfair due to cultural and language differences in the test takers. As Airasian (1997) noted, our society places emphasis on equality for all students. Cultural and language differences are viewed positively rather than indicating intellectual inferiority. Airasian also stated that, although certain intelligence tests still play an important role in the psychological fields, in education the intelligence test is no longer viewed as a key tool for ranking and evaluation.

Gatekeeper Courses

Introductory quantitative courses are often the gateway to science, mathematics, engineering, technology (SMET), and to business majors courses. Calculus, chemistry, or accounting do not come easily for many students interested in these fields of study. They frequently fail to complete these courses satisfactorily on their first attempt (Andrade, 1999; Gainen, 1995; Seese, 1994). In her article, Gainan (1995) listed several sources, including the National Academy of Sciences, which estimated that 35% of students who choose SMET majors when they enter college switch out of them or leave college altogether between their freshman and sophomore years. Further documentation is presented that the most dramatic and frequent limiting factor is insufficient or inadequate preparation in mathematics. Mathematics ability is a strong predictor of success in engineering and is highly correlated with GPA in accounting. Thus, the introductory mathematics and science courses take on a gatekeeper role for progression into SMET degree programs (Andrade, 1999). Students who do not pass these gatekeeper courses are often judged as lacking the analytical ability when in fact it may be a lack in preparation in mathematics or science.

Calculus I is the mathematics gatekeeper course for SMET majors. Over one half million students have enrolled in calculus in four-year colleges and universities annually and half of these students have enrolled in mainstream engineering calculus (Ferrini-Mundy & Graham, 1991). Multi-sections of this course are often taught to accommodate these students interested in the SMET fields (Seese, 1994). In a department where multi-section courses are taught by professors, instructors, part-time teachers, and graduate assistants, a problem with consistency could exist. At the college level, a long history of academic freedom exists, and professors and instructors expect to be able to individualize their teaching process and assessment methods. Generally, a department committee defines the basic content of the course, but the approach to teaching and assessment methods invariably are quite different.

Evaluating students in multi-section gatekeeper courses is problematic. If a student receives a B in one section of a course, the question becomes whether that student has mastered the knowledge and skills of the material to the same level as another student who received a B in a different section of the same course. Andrade (1999) wrote that a high failure rate in a class may imply high academic standards or it may indicate curricular and instructional problems. Conversely, a large number of passing grades may mean lowered standards, or could reflect a more coherent curriculum or improved instructional strategies. Passing Calculus I does not necessarily assure success in the following courses in the calculus sequence. One convincing measure of success would be the students' grades in the next course or courses in the curricular sequence. For a gatekeeper course to be judged effective, the majority of students who successfully completed it should be able to pass the next course on the first try. When multiple sections of a course are offered, Seese (1994) pointed out the significance of consistency in the course. Results of a survey conducted by Cizek, Fitzgerald, and Rachor (1995) indicated that teachers' assessment practices were highly variable and unpredictable from characteristics such as gender, years of experience, grade level, familiarity with assessment policies and practice settings such as frequency of grading, methods of combining marks and types of marks used. Therefore, consistency in the course should also include consistency in methods of evaluating students on their knowledge and skill levels. The problem in this situation is to ensure that faculty can retain control over their own courses and at the same time ensure that all students who pass the gatekeeper course will be properly prepared to be successful in follow up courses. The question is all about a balance between common academic outcomes (measured by common course finals) and academic freedom of instruction (Duckwall & Wilson, 1996).

ACT Assessment Program

The ACT (American College Test) Assessment Program was established in 1959 as a not-for profit organization and has grown to become one of the leading providers of educational assessment and research services in the world. The program was established by E. F. Lindquist who believed that the test should measure what students can do with what they have learned throughout their elementary and high school years. His philosophy has guided the specifications for the ACT Assessment educational achievement tests to ensure that what is being assessed reflects what is being taught in classrooms nationwide and what is necessary for college success (American College Testing Program, 1996).

The ACT is published and graded by the American College Testing Program of Iowa City, Iowa. The SAT, a competitor of the ACT, is governed by the College Board in Princeton, New Jersey. The country is divided by state in their usage, with the middle states using the ACT and the eastern and western states using the SAT. The prime purpose of both the ACT and the SAT is to assist in predicting college success (Bontekoe, 1992).

Every year hundreds of thousands of high school juniors and seniors take the ACT test to assist them in planning their educational and career goals as they prepare to make the transition from high school to college (Morgan, 1992). During the academic year 2001-2002 over 2 million ACT Assessments were administered (Act: News: Facts about the ACT assessment, Retrieved February 7, 2003). Colleges have used the scores as one of three major pieces of data, along with high school grade point average (GPA) and rank in class, to help decide if an applicant meets entrance requirements, course placement, and as a measure for awarding scholarship dollars (Bontekoe, 1992).

The ACT Assessment, or "A-C-T" as it is commonly called, is an examination that covers four skill areas: English, mathematics, reading, and science reasoning. The ACT includes 215 multiple-choice questions and takes approximately 3 hours to complete. Students receive 12 scores on the ACT Assessment – four skill area test scores, seven sub-scores, and one composite score (Morgan, 1992).

The ACT Assessment Mathematics Test is designed to assess the mathematical skills that students have typically acquired in courses taken up to the beginning of grade 12 and that are relevant to success in college. Table 2.1 describes the areas of

Table 2.1

ACT Assessment Mathematics Test Content Specifications

Content Area	Number and Percent of Questions	
Pre-Algebra	14	(23%)
Elementary Algebra	10	(17%)
Intermediate Algebra	9	(15%)
Coordinate Geometry	9	(15%)
Plane Geometry	14	(23%)
Trigonometry	4	(7 %)
Total Questions	60	(100%)

mathematics that are tested. Scores on the four test areas and the composite score can range from 1 (lowest) to 36 (highest). This scale is the key for decision making for high school counselors, students, parents and college admissions offices.

ACT scores can be interpreted by means of describing specific academic skills and knowledge that a student has likely acquired, given the particular score range in which her or his ACT score lies. ACT's Standards for Transition offers this criterion-referenced type of score interpretation. Since the ACT Assessment is a curriculum-based assessment designed to measure high school students' general educational development, the Standards for Transition have been written to show how skills can progress. These Standards for Transition are comprised of six score ranges: 13-15,16-19, 20-23, 24-27, 28-32, and 33-36. The same ranges are used in all four subject areas, where each range has a list of statements that describe what a student is likely to know and to be able to do in that area (see Appendix A for the set of Mathematics Standards for Transition for each score range). The Standards for Transition are cumulative: that is, students in a given score range can typically demonstrate most or all of the knowledge and skills in the preceding score ranges. These ranges were developed by content area experts and nationally recognized scholars from high schools and universities who determined the knowledge and skills needed to correctly answer test questions corresponding to various score ranges. The Standards give insight into the meaning of a score, and they serve as a direct link between what students have learned and what they are ready to learn next (ACT: Standards for Transition, Retrieved March 4, 2003). The purpose of these Standards for Transition is to help better understand how the scores relate to the kinds of skills needed for success in college, and in particular for course placement. They can also be used to help teachers and other educators learn more about their students' academic strengths and weaknesses (ACT: Research: Information Brief:2002-2, Retrieved March 4, 2003).

A crucial issue is whether the ACT is a legitimate criterion for making admissions and scholarship decisions at a particular institution. Reports by Bontekoe (1992) and Morgan (1992) listed a number of studies on the ACT which were not very supportive of its use as a predictor of success at the college level, yet also listed many extensive studies that did support the use of ACT scores to be valid predictors. Bentekoe's (1992) own study at Trinity Christian College in Palos Heights (Illinois) found that college GPA could be reasonably predicted from the six defined ACT Composite ranges. The correlation between high school GPA and college GPA was even higher than that between ACT scores and college GPA. A study by Morgan (1992) found that there was a statistically significant, but low, correlation between ACT composite scores and GPA of college freshmen. Her findings also suggested that other intervening factors such as high school rank and high school GPA could have a predictive influence on college GPA. In her report, Morgan also found the research on the relationship of ACT composite scores as predictors of success in college to be limited. In a relational study by Levy (2001) conducted in three Mississippi Community Colleges, findings indicated that when combined, high school grade point average and ACT mathematics sub-scores were strong predictors of college mathematics grade point average and even stronger predictors of developmental students' total mathematics grade point average.

Critics argue that standardized tests, such as the ACT, are culturally and linguistically biased. Thus, in 1989 a major change in organization and test item selection by the ACT program set out to end racial, cultural, and gender bias in the test and to more closely address the needs of college decision makers (Bontekoe, 1992). Other critics feel that students are over-tested, and that colleges place too much emphasis on the work of one morning in a pressure-filled testing center for a test to be a reliable predictor of success in college (Morgan, 1992). Those in defense of the test feel it is the only objective measure of a student's academic abilities because the subjective grading of teachers, the widely divergent academic expectations of high schools in widely different communities, and the inconsistency of an individual's work habits have been eliminated as interfering factors. In other words, tests may be flawed but still useful.

A report by the American College Testing Program (1996) described the steps ACT has taken to ensure that the content foundations of the ACT Assessment mathematics test are solid. The report described how the content and skills measured are determined, and how they are kept up-to-date. Survey results were included to verify that the test is consistent with what is being taught in the schools. Reviews of state and national curriculum and performance standards were presented that support the validity of the test at the high school level. There was documentation of the role U.S. mathematics teachers play in developing the ACT mathematics test and finally results of studies that presented the mathematics test to be effective in measuring the mathematical skills important for college success.

This Study

Most colleges and universities offer a calculus sequence of studies through their mathematics department. These sequences may be of various formats. The most common formats may be the four semester sequence consisting of Calculus I, II, III and IV with each earning three credit hours, or the three semester sequence, Calculus I, II and III each earning four semester hours. One theme all variations have in common is that Calculus I is the gatekeeper course of the sequence. Many studies have been conducted in the area of teaching and learning of the calculus concepts, but in the area of assessment of calculus as a gatekeeper course, the literature is limited and the studies are practically nonexistent. Some studies were found outside the mathematics area on common final examinations, however they were mostly studies done on the content of the examination, but none were found comparing the presence or absence of the examination itself.

Some studies have been conducted in related areas. At Miami Dade Community College in Florida, a study by Bashford (2000) suggested that passing prerequisite courses does not ensure that students are adequately prepared for the next mathematics course in the sequence. In most cases, students placed directly into courses by scores on a placement test were more successful than students who progressed through college preparatory or prerequisite courses. In Johnson County Community College in Kansas, a study was conducted to see if core components in mathematics courses could replace the comprehensive common course final examination (Duckwall & Wilson, 1996). This study was conducted in an effort to balance common academic outcomes with faculty's individual teaching styles. The core components were designed by faculty who agreed to grade them identically and include assessment of them as part of the course final examination. Results indicated a positive correlation between student scores on the components and course grades as a whole; however, issues arose in possible bias in scoring, potential for grade inflation and concerns about the core components' ability to measure final exam objectives.

At the University of California Santa Barbara, Fisher (1996) conducted a study in 1995 on the validity of multiple choice and performance-based testing as a predictor of undergraduate mathematics and chemistry achievement. Results indicated that performance-based tests had statistically significant correlations with grades in algebra, calculus for the physical sciences, calculus for the social sciences, and chemistry. Multiple choice testing of pre-calculus concepts had a higher correlation than performance-based testing, but a combination of both provided the best test to predict mathematics or chemistry achievement.

A study was conducted on entering freshmen mathematics students' placement into courses at West Virginia University. Ahrens (1980) found that the two best predictor variables were high school background and the grade on a departmental placement examination. Other variables used were the student's ACT scores and advisor recommendation.

These few studies on assessment in the field of mathematics examine a very broad scope of research questions, each of which is somewhat related to the current study. This study focused on the assessment practices of a Calculus I course, a prerequisite to the remainder courses in the sequence. The course is often offered via multi-sections causing concern for consistency in evaluation methods.

Summary

The study of calculus is an integral part of the Science, Mathematics, Engineering and Technology disciplines. Calculus I is the gatekeeper course for subsequent Calculus courses. With many sections of Calculus I offered at a major institution, consistency in student preparation may become an issue. A standardized departmental final examination at the end of Calculus I has often been used to ensure that a student is appropriately prepared to advance to the next level. This examination may come in varying formats. Most often a multiple-choice format is used since it can be administered efficiently and can easily be graded when a large number of students are involved. Other formats that are sometimes used are the open-ended essay question or an oral response type of examination. Whether this examination is an effective tool for monitoring the level of preparation of the student in a Calculus I course is the main focus of this research.

The literature on this topic is virtually non-existent. A few studies have been found in closely related areas, but none that related the effect of the presence or absence of standardized multiple-choice departmental final examination in a multi-course sequence.

CHAPTER III

METHODS

This chapter outlines the design of this research study. An explanation of the research design is presented. The sample and sampling methods are described. The data collection procedures of the study and the method of analysis of the data are outlined. Also, the limitations, reliability, and validity of the study are discussed.

Research Design

The design of this research is causal-comparative. The causal-comparative method is an approach to exploring cause-and-effect relationships between phenomena. It is a type of research that seeks to determine possible cause and effect of a behavior pattern or personal characteristic by comparing individuals in whom it is present with individuals in whom it is absent or present to a lesser degree (Gall, Borg & Gall, 1996). The main reason for using the causal-comparative method is that the cause-and-effect relationship in this study is not open to experimental manipulation. The causative factor is the presence or absence of the standardized multiple-choice departmental final (SMCDF) examination for Calculus I students and the effect is the students' level of success in Calculus II through Calculus IV as measured by the final grades in the course. The researcher also investigated students' ACT mathematics scores as a possible relating factor that might explain any observed differences between the two groups.

Sample

The sample used in this study was obtained from a population of students who completed Calculus I at Mississippi State University (MSU) from the fall of 1995 through the spring of 1999 semesters. MSU is a land-grant institution and is a comprehensive, doctoral-degree-granting university housing 52 academic departments in eight colleges and schools including the College of Engineering, the College of Veterinary Medicine, and the School of Architecture. MSU is one of the largest of the eight Mississippi public universities with enrollment of around 16,000 students. About 75% of the student population is from the state of Mississippi, 19.9% is from out-of-state, and 5.1% are international, representing 81 countries. The university facts and figures indicate around 20.3% of its population are minority students while 54% of the university population are male students. Mississippi State University is the fourth largest industry in the state of Mississippi (MSU General Facts and Figures, Retrieved March 16, 2003).

Each semester multiple sections of MA 1713 (Calculus I) through MA 2743 (Calculus IV) are taught by professors, part-time teachers, or graduate students. The common departmental examination group (DEPT) in the study originated from a total of 28 sections of Calculus I taught during the fall 1995, spring 1996, fall 1996, and spring 1997 semesters (see Table 3.1). These Calculus I students were required to take the Calculus I standardized multiple-choice departmental final (SMCDF) examination as part of the requirement of the course. The instructor examination (INST) group in the study originated from a total of 32 sections of Calculus I student grades from the fall 1997, spring 1998, fall 1998, and spring 1999 semesters (see Table 3.1). These Calculus I

Table 3.1

Number of Sections and Teachers in Calculus I-IV Included in Study

DEPT	* I	II	III	IV
Fall 95	Δ(9/7)			
Spring 96	(4/4)	(7/5)		
Fall 96	(10/7)	(4/3)	(5/5)	
Spring 97	(5/5)	(7/6)	(2/2)	(4/4)
Fall 97		(4/3)	(5/5)	(3/3)
Spring 98			(4/4)	(4/4)
Fall 98				(3/3)
Total Sections	(28/23)	(22/17)	(16/16)	(14/14)
INST	**I**	**II**	**III**	**IV**
Fall 97	(11/9)			
Spring 98	(5/3)	(7/5)		
Fall 98	(11/7)	(4/4)	(5/5)	
Spring 99	(5/4)	(7/5)	(4/4)	(4/4)
Fall 99		(4/4)	(5/5)	(3/3)
Spring 2000			(3/3)	(4/4)
Fall 2000				(3/3)
Total Sections	(32/23)	(22/18)	(17/17)	(14/14)

*I =MA 1713 Calculus I　　II =MA 1723 Calculus II
III =MA 2733 Calculus III　　IV =MA 2743 Calculus IV
$\Delta(9, 7) \rightarrow$ (9 sections and 7 teachers)

students were not required to take the Calculus I SMCDF examination since the SMCDF examination was discontinued beginning in the fall of 1997. Instead, these Calculus I students were required to take their instructor generated final examination. The professors, instructors or graduate students teaching the classes of both groups were required to cover material outlined in a syllabus which was developed by a committee from the department of mathematics. The department further required a final examination at the end of the course. All sections taught used the same textbook. The department calculus committee also created a syllabus for Calculus II, III and IV, however a final examination was not departmentalized.

Summer classes were not included in the study because the use of the departmental final examination was not required during the years when the SMCDF examination was used. It was left up to the discretion of the individual instructor whether to use an old departmental final examination or use a teacher-generated final examination during all summer sessions.

Most classes began with approximately 45 students in each section. From the 28 sections of Calculus I of the DEPT group population, only students who completed the course with a grade of A, B or C and then completed Calculus II at (MSU) were included in the Calculus II-DEPT group (see Table 3.2). From the 32 sections of Calculus I of the INST group population, those earning a grade of A, B or C and then completed Calculus II at MSU were assigned to the Calculus II-INST group (see Table 3.2). The INST group

Table 3.2

DEPT and INST Samples

Calculus II, III, IV-DEPT	Calculus II, III, IV-INST
F95-I → II → III → IV	F97-I → II → III → IV
S96-I → II → III → IV	S98-I → II → III → IV
F96-I → II → III → IV	F98-I → II → III → IV
S97-I → II → III → IV	S99-I → II → III → IV

grades reflect teacher made final examinations. Only student grades of A, B or C in Calculus I were selected for the sample since students with D and F grades had to repeat the course to obtain credit.

The researcher further investigated if there was a lasting effect in any existing differences in the level of success between the two groups in subsequent calculus courses. For this reason, student grades were also matched to Calculus III and Calculus IV grades. The Mathematics Department did not create a common final examination to be used in Calculus II, III or IV; instead, these courses had teacher-generated final examinations.

Students from the Calculus II-DEPT and Calculus II-INST groups who passed Calculus II with a grade of C or better were matched with their Calculus III grades if the course was taken at MSU. This created the Calculus III-DEPT and Calculus III-INST groups for statistical testing (see Table 3.2). Furthermore, students from the Calculus III-DEPT and Calculus III-INST groups who passed Calculus III with a grade of C or better were matched with their Calculus IV grades if the course was taken at MSU. This created the Calculus IV-DEPT and Calculus IV- INST groups (see Table 3.2). The following describes the relationship between the groups (⊂ indicates subset):

Calculus IV-DEPT ⊂ Calculus III-DEPT ⊂ Calculus II-DEPT ⊂ DEPT-I

Calculus IV-INST ⊂ Calculus III-INST ⊂ Calculus II-INST ⊂ INST-I

To investigate the effect of ACT Mathematics scores on any of the results between the DEPT and INST groups, the researcher sorted the groups using the characteristic of the ACT Mathematics Standards for Transition score ranges (Appendix A).

Procedures

The researcher received approval to conduct the study from the Institutional Review Board (IRB) at Mississippi State University (MSU) (see Appendix B). Data were obtained from the Office of Information and Technology Services (ITS) at MSU with permission of the registrar's office. In order to preserve anonymity of the students, names and Social Security numbers were removed from the data before release to researcher. A separate data set was requested for each semester. Each student record on each data set was numbered consecutively starting with the number one. Each record in the data set contained an ACT mathematics assessment score, Calculus I grade and term attended, a possible Calculus II grade and term attended at MSU, a possible Calculus III grade and term attended at MSU and a possible Calculus IV grade and term attended at MSU. Grades were assigned quality points as follows:

A = 4
B = 3
C = 2
D = 1
F = 0

The Departmental Final Examination

Each semester students are required by the department to take some type of final examination at the end of their Calculus I course. From 1983 until the spring 1997 semester, this final examination was comprehensive, standardized, in a multiple-choice format and created by a committee in the MSU Department of Mathematics and Statistics. There were 40 multiple-choice questions on this examination (see Appendix C for a sample examination). It was administered to all Calculus I students at the same time and graded electronically by the Office Institutional Research at MSU. Beginning in the fall of 1997 this final examination policy was changed. A final examination was still required, but it was no longer standardized and written by a committee from the mathematics department. The examination was then teacher-generated and graded, and administered during the regular final examination schedule. For both situations, a student's final course grade was determined according to a departmental policy. The policy stated that the final examination counted as 50% of the student grade if the student's examination grade was equal to or higher than the student's class average up to the time of the final exam. If the student's final examination grade was lower than the student's class average the final examination would count as one third of the course grade (Department of Mathematics and Statistics, personal communication, March 2003).

The Null Hypotheses

The purpose of this study was to determine if there is a difference in students' success in subsequent Calculus courses before and after use of the standardized multiple-choice departmental final (SMCDF) examination for Calculus I. The null hypotheses for the study are stated as follows:

HO_1: There will be no statistically significant difference in the mean Calculus II course grades between Calculus II students who have taken a Calculus I SMCDF examination (Calculus II-DEPT) and Calculus II students who have not taken a Calculus I SMCDF examination (Calculus II-INST).

A two-way analysis of variance (ANOVA) was calculated to compare the means of the Calculus II grades of the DEPT and the INST groups while controlling for the influence of ACT-mathematics scores.

HO_2: There will be no statistically significant difference in the mean Calculus III course grades between Calculus III students who have taken a Calculus I SMCDF examination (Calculus III-DEPT) and Calculus III students who have not taken a Calculus I SMCDF examination (Calculus II-INST).

A two-way analysis of variance (ANOVA) was calculated to compare the means of the Calculus III grades of the DEPT and the INST groups while controlling for the influence of ACT-mathematics scores.

HO_3: There will be no statistically significant difference in the mean Calculus IV course grades between Calculus IV students who have taken a Calculus I SMCDF examination (Calculus IV-DEPT) and Calculus IV students who have not taken a Calculus I SMCDF examination (Calculus IV-INST).

A two-way analysis of variance (ANOVA) was calculated to compare the means of the Calculus IV grades of the DEPT and the INST groups while controlling for the influence of ACT-mathematics scores.

HO_4: There will be no statistically significant differences in the mean Calculus II, III, and IV course grades of students having an ACT mathematics assessment score of 16-19 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

A two-way analysis of variance (ANOVA) was run to compare the means of the Calculus II, III, IV grades between the DEPT and the INST groups at the 16-19 ACT-mathematics Standards for Transition range

HO_5: There will be no statistically significant differences in the mean Calculus II, III, and IV course grades of students having an ACT mathematics assessment score of 20-23 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

A two-way analysis of variance (ANOVA) was calculated to compare the means of the Calculus II, III, IV grades between the DEPT and the INST groups at the 20-23 ACT- mathematics Standards for Transition range.

HO_6: There will be no statistically significant differences in the mean Calculus II, III, and IV grades of students having an ACT mathematics assessment score of 24-27 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

A two-way analysis of variance (ANOVA) was calculated to compare the means of the Calculus II, III, IV grades between the DEPT and the INST groups at the 24-27 ACT- mathematics Standards for Transition range.

HO_7: There will be no statistically significant differences in the mean Calculus II, III, and IV grades of students having an ACT mathematics assessment score of 28-32 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

A two-way analysis of variance (ANOVA) was calculated to compare the means of the Calculus II, III, IV grades between the DEPT and the INST groups at the 28-32 ACT- mathematics Standards for Transition range.

HO_8: There will be no statistically significant differences in the mean Calculus II, III, and IV grades of students having an ACT mathematics assessment score of 33-36 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

A two-way analysis of variance (ANOVA) was calculated to compare the means of the Calculus II, III, IV grades between the DEPT and the INST groups at the 33-36 ACT- mathematics Standards for Transition range.

The Statistical Analysis

In this section the statistical procedures used to analyze the data in the study are described. The independent variable group had two levels, Calculus I-DEPT and Calculus I-INST. The dependent variable was scores, measured at all three subsequent levels of calculus, Calculus II, Calculus III and Calculus IV. To further investigate the research question, student ACT Mathematics scores were converted to ACT Mathematics Standards for Transition score ranges, and were also used as an independent variable. There were 6 ranges: 13-15, 16-19, 20-23, 24-27, 28-32, and 33-36. Again the dependent variable was scores, measured in the Calculus II, Calculus III, and Calculus IV courses. The Statistical Package for the Social Sciences for the PC (SPSS-PC) was used in the statistical analysis.

For each comparison group in the study, the group means and standard deviations were first calculated. Because the researcher was interested in determining if the mean scores between the two groups (DEPT and INST) differed significantly and whether the ACT Mathematics scores had any effect on any of the differences, a two-way analysis of variance (ANOVA) was used. Since the ANOVA uses two categorical independent variables, ACT Mathematics scores were summarized into Standards for Transition score ranges as one independent variable and the other was the two groups (DEPT and INST). Assumptions of normality of data was established by checking for the skewness and kurtosis of the data sets, while homogeneity of variance was checked using Levene's Test of Equality of Error Variances.

Once the assumptions were verified, the researcher examined whether the two samples at the three levels of Calculus being compared varied systematically with their scores along the characteristic of the predefined ACT Mathematics Standards for Transition score ranges.

Limitations

The data in this study spanned the extended period of time from1995 to 2001. During this time frame, there were numerous changes in the Calculus courses that might have affected the results of this study. Some of the key changes had been in the use of technology in the classrooms by both teachers and students. In the fall of 1995 the use of the graphing calculator had become a requirement for all Calculus I classes at MSU and the end of the semester final examination was made graphing calculator sensitive (Neumann, 1995). Teachers, some with little background with the graphing calculator, had improved in their ability to utilize graphing calculators and computer algebra systems in their teaching methods. Students had also increased their technology skills.

Some school policies could have affected the results of the study. From the fall of 1992 to the spring of 1995, MSU had a policy that allowed students to use up to five drops, called super drops, during their college career. Super drops allowed the student to drop the class any time up to the last day of withdrawal listed in the semester bulletin and receive a W grade on their transcript. In the fall of 1995 this policy was changed to super audits, where the student could change to an audit any time up to the last day of withdrawal stated in the semester bulletin (Registrar's Office Tool Kit, 1995). The teacher was to submit an F on the record only if the student stopped coming to the class (B. Stokes, MSU registrar's office, personal communication, March 7, 2003). The super audit was eliminated by the fall of 1997, at which time the current policy was set that allowed a student to drop a course with a W grade after the tenth class day through the 30th class day into the semester with the approval of his/her advisor. A student cannot drop courses after this period except in documented cases of extreme situations (MSU, 2001).

An obvious limitation to the study was the many different professors, instructors and graduate students teaching the calculus courses. Each had their own teaching styles, methods of assessment, experience in teaching, and expertise in the material taught.

Furthermore, each semester the instructors appointed to teach each of the calculus courses changed, bringing about more inconsistencies.

One fact that needs to be pointed out was that students in some majors such as computer science and the biological sciences were not required to take the upper level courses in the calculus sequence, whereas engineering and physics majors had to take all four courses. Information was not available as to which records represented these students.

Finally, textbooks changed during the time period of this study. Textbook changes were made to accommodate technological advancements and curriculum reform. Table 3.3 lists the textbooks that were using during the time period of the study. These changes could have an impact on the student grades thus must be considered as a limitation.

Table 3.3

Change in Textbooks

Semester	Title	Edition	Author	(Year)
Fall 1995	Calculus	3rd	Berkey-Blanchard	(1995)
Fall 1998	Concepts and Context	1st	James Stewart	(1998)
Spring 1999	Calculus	3rd	James Stewart	(1999)
Fall 2000	Calculus	4th	James Stewart	(2000)

Reliability and Validity

Gall et al. (1996) described internal validity as the extent to which extraneous variables can be controlled by the researcher so that any observed effect can be attributed solely to the treatment variable. Reliability, as applied to educational measurements, is the level of internal consistency or stability of the measuring device over time. An instrument can be reliable (able to be repeated with consistency), but not valid; however, if an instrument is valid, it must also be reliable.

This study measured student scores over three consecutive semesters. For this length of time many unpredictable or extraneous variables could have occurred causing an effect on the student scores. Some students might have matured earlier than others during this period of time, while some students might have had a dramatic occurrence in their personal lives that could have affected their learning. These types of variables cannot be quantified or controlled. The ACT Mathematics score, however, was a variable that was quantified and investigated. Since the ACT Mathematics score allowed for students to be categorized by ability level for statistical testing, it presented an opportunity for adding validity to this study. The American College Testing Program, which produces the ACT Mathematics test, conducts regular research on their tests' reliability and validity (ACT Research, Retrieved February 7, 2003). Their Standards for Transition ranges are clearly defined and thus serve as a valid instrument that produces reliable scores for measurement purposes.

Quality points ranging from 0 to 4 were used in this study for student scores. A 4 represents an A, a 3 is considered a B, a 2 is a C, a 1 is a D, and 0 is an F (Bulletin of

Mississippi State University, 1997). Letter grades in the Calculus courses were determined on a ten point scale where A was assigned to a percentage grade ranging from 90% to 100%, B came from a percentage grade of 80% to 89%, C from 70% to 79%, D from 60% to 69% and F represented any score below 60%. There is considerable difference between the high and low end of the 10 point scale of the percentage score causing concern for validity and reliability; however, a large sample size for each group would bring more reliability to the data. No reliability or validity measures were applied to these scores in this study.

CHAPTER IV

RESULTS

This chapter presents the findings of the research conducted to determine if there is a difference in student's success in the calculus sequence when a standardized multiple-choice departmental final (SMCDF) examination is administered at the end of the first calculus course. First, the purpose of this study is reiterated. This follows with an overview of the data obtained for the study where grades and ACT scores are organized for the groups in question. The next section describes the statistical procedures used to examine the research questions under investigation and results of these procedures are reported. Finally, a summary of the results will conclude this Chapter.

Purpose

The purpose of this study is to determine if there is a difference in students' success in subsequent calculus courses before and after use of the standardized multiple-choice departmental final (SMCDF) examination at the end of the Calculus I course. If there is a difference, is this difference consistent along varying levels of student's ACT-Mathematics assessment scores. Mean scores of student grades over a period of 4 semesters when the SMCDF examination was in effect were compared to mean scores of student grades over a period of 4 semesters when the test was no longer in effect.

Descriptive Data

Data were obtained from the Mississippi State University Department of Information and Technology Services. The data consisted of all Calculus I student records from the fall of 1995 through the spring of 1997 semesters for the DEPT (departmental final examination) group and the fall 1997 through the spring of 1999 semesters for the INST (instructor final examination) group, excluding summer sessions. Following IRB approval (see Appendix B), data were received in Excel spreadsheet format with each record containing the student's ACT composite score, ACT-Mathematics score, MA 1713 (Calculus I) grade, MA 1723 (Calculus II) grade, MA 2733 (Calculus III), and MA 2743 (Calculus IV) with each respective term attended. If a student never registered for a subsequent course the cell was left blank and later converted to "Does Not Exist" or DNE. An S, representing a satisfactory grade, reported students who entered MSU for the first time with a passing score on the Advanced Placement Examination taken in their high school years. AU represented an audit and W identified a withdrawal without penalty from the course (Bulletin of Mississippi State University, 1997). These records were eliminated from the study prior to statistical testing. For a complete set of tables of frequencies and percents of all grades and ACT-Mathematics scores obtained from ITS, see Appendix D. For statistical testing purposes all letter grades were converted to Quality Points where A=4, B=3, C=2, D=1, and F=0.

The DEPT group began with a total of 1063 records, of which 944 included ACT-Mathematics scores. The INST group began with a total of 1497 records, of which 1383 included ACT-Mathematics scores. Figure 4.1 shows the frequency of the different grades in both groups.

Calculus I

	DEPT	INST
A	135	221
B	219	222
C	216	264
D	109	162
F	91	299
AU	193	2
W	28	145
S	80	182
TOTAL	1063	1497

Calculus II

	DEPT	INST
A	87	96
B	95	100
C	88	124
D	29	53
F	39	85
AU	83	1
W	14	40
DNE	135	208
TOTAL	570	707

Calculus III

	DEPT	INST
A	41	70
B	56	75
C	56	38
D	19	19
F	15	25
AU	28	0
W	8	14
DNE	47	79
TOTAL	270	320

Calculus IV

	DEPT	INST
A	52	58
B	49	50
C	13	34
D	8	12
F	5	6
AU	1	0
W	4	5
DNE	21	18
TOTAL	153	183

Figure 4.1: Frequency of Grades in DEPT and INST Groups

Since the two groups did not have the same number of records, a clearer comparison was possible by calculating percentages for the different grades in both groups at all four levels of calculus (see Table 4.1). A graphical representation of these percentages of grades A through F for Calculus I, II, III and IV can be found in Appendix E. While the DEPT group showed a fairly normal distribution of grades for Calculus I with concentration in the B's and C's, the INST group showed the largest count or 26% of grades in the F, or failing mark as compared to the DEPT group which showed 12% failure grades in Calculus I. In Calculus II the majority of grades for both the DEPT and INST groups lay in the B's and C's even though the INST group still showed a fairly large number or 19% of the grades as failing. For Calculus III the DEPT group still had its concentration in the B's and C's, whereas, the Calculus III INST group had mostly B's, followed by A's. In Calculus IV both groups had a small number of low grades and a high concentration of A's followed by B's.

Table 4.1

Grade Distributions Using Percents for DEPT and INST Groups

Grade:	A	B	C	D	F
Calculus I					
DEPT	18%	29%	28%	13%	12%
INST	19%	19%	22%	14%	26%
Calculus II					
DEPT	26%	28%	26%	9%	11%
INST	21%	22%	26%	12%	19%
Calculus III					
DEPT	22%	30%	30%	10%	8%
INST	31%	33%	17%	8%	11%
Calculus IV					
DEPT	41%	39%	10%	6%	4%
INST	36%	31%	21%	8%	4%

After tabulating the number of records that belonged in each of the ranges by Standards for Transition, it was found that only 6 students from the DEPT group and 11 students from the INST group were in the “less than or equal to 15” ACT-Mathematics transition range. Since there were so few at this range, these student grades were combined with the next transition range of “16 to 19” and renamed “less than or equal to 19”. This reduced the six transition ranges to five. Table 4.2 shows the frequencies and percents of the ACT-Mathematics score ranges by Standards for Transition for the DEPT and INST groups. Both the DEPT and INST groups found over 50% of the students in the 24-27 and 28-32 ACT-Mathematics ranges. The missing System represents records with unreported ACT-Mathematics scores.

Table 4.2

Frequency and Percent of ACT-Mathematics Scores by Range Levels of DEPT and INST Groups of the Original Data Set

	Frequency		Valid Percent		Cumulative Percent	
	DEPT	INST	DEPT	INST	DEPT	INST
ACTm<=19	107	170	11.3	12.3	11.3	12.3
ACTm=20-23	215	272	22.8	19.7	34.1	32.0
ACTm=24-27	300	410	31.8	29.6	65.9	61.6
ACTm=28-32	286	469	30.3	33.9	96.2	95.5
ACTm=33-36	36	62	3.8	4.5	100.0	100.0
Total	944	1383	100.0	100.0		
Missing System	119	114				
Total	1063	1497				

According to the MSU Bulletin (2002), a grade of C or higher is a prerequisite for advancing to the next course in the sequence of the calculus classes. For this reason, only records of passing Calculus I grades were valid records to test for significant differences in grades between Calculus II-DEPT and Calculus II-INST. Similarly, only records of passing Calculus II grades were used to test for significant differences in grades between Calculus III-DEPT and Calculus III-INST and records of passing Calculus III grades were used to test for significant differences in grades between Calculus IV-DEPT and Calculus IV-INST. Figure 4.2 gives an overview of the number of records used (N-value) in the study at each subsequent level for the DEPT and INST groups in the study. When ACT-Mathematics scores are incorporated, the N-value was reduced due to records with missing ACT scores. These new N values are reported in parenthesis.

Table 4.3 presents the percentages of students that passed each level of calculus. S grades were not included in the Calculus I group since these students took the course as an Advanced Placement (AP) course in high school and did not take Calculus I at MSU.

Table 4.3

Percents of Students Passing Calculus

	Group	Passing	Total*	Percent
Calculus I	DEPT	570	983	57.99%
	INST	707	1315	53.76%
Calculus II	DEPT	270	435	62.07%
	INST	320	499	64.13%
Calculus III	DEPT	153	223	68.61%
	INST	183	241	75.93%
Calculus IV	DEPT	114	132	86.36%
	INST	142	165	86.06%

Total = Number of calculus records less S (satisfactory) records and DNE (non existent) records.

Students with an AU had to be counted since they registered and attended the course for at least a minimum amount of time. Table 4.3 shows that the DEPT group of students who took Calculus I and were required to take the SMCDF examination at the

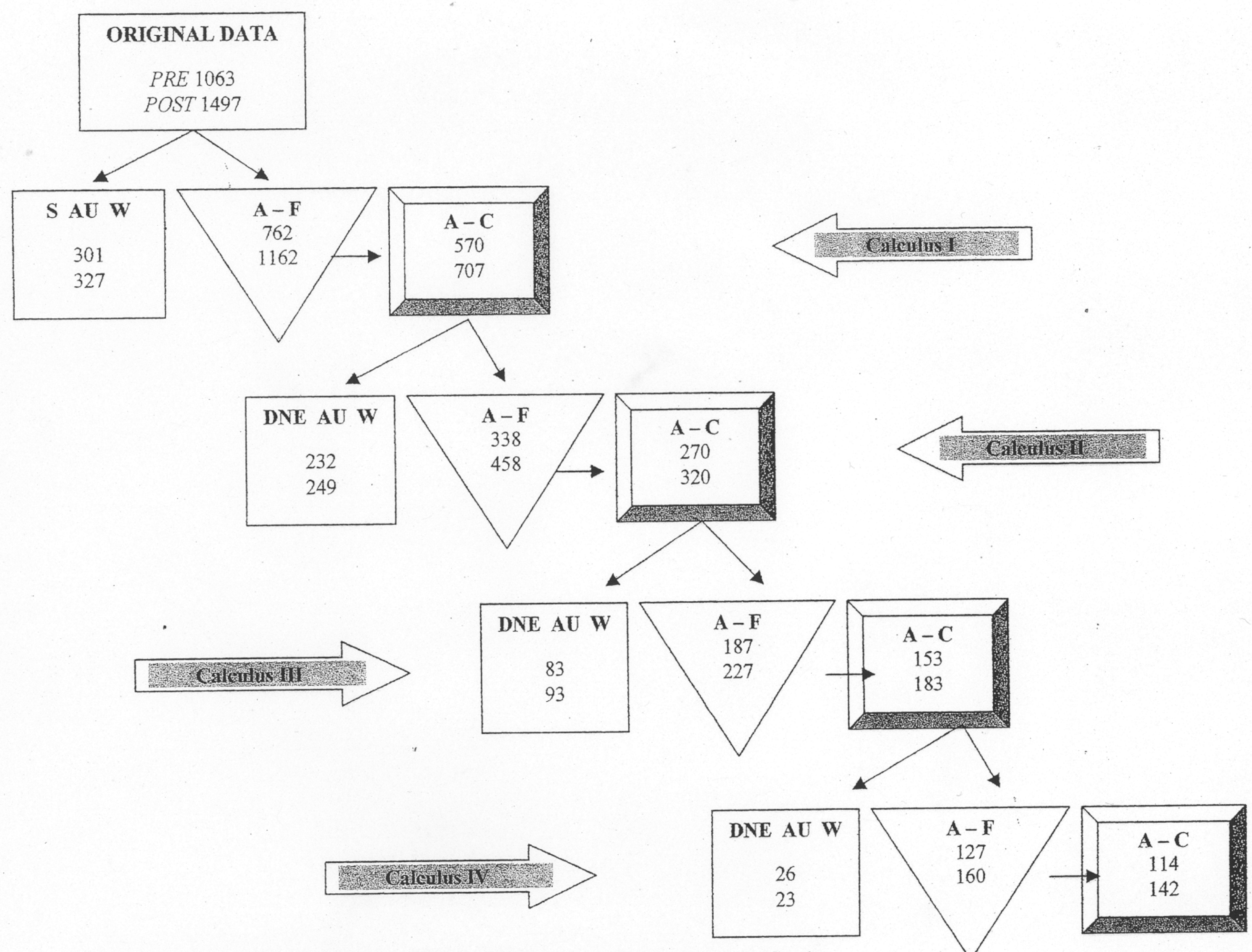


Figure 4.2: **Number of Records (N-value) of DEPT and INST Groups at Subsequent Calculus Levels**

end of the course had about 4% higher passing rate as compared to the INST group where the examination was not required. When these students were followed into Calculus II, the percentages switched and the INST group showed about a 2% higher passing rate, following by the same group having a 7.5% higher passing rate in Calculus III. By the time these passing students reached Calculus IV the percentages were about even for the two groups. One final observation to be noticed from Table 4.3 is that the DEPT group had an 11.6% rate of students completing the entire calculus sequence (114 passing Calculus IV / 983 total starting in Calculus I) and the INST group had a 10.8 % rate of completion (142 passing Calculus IV / 1315 total starting in Calculus I). Mention must be made that not all students were required to take all four courses in the sequence for their major area of study, thus these percents do not indicate necessarily that students are failing out of the course sequence. Information was not available for making the distinction between the different cases.

Data Analysis

Data were analyzed using SPSS for Windows, Release 11.0. The purpose of this study was to determine if there was a difference in students' success in subsequent calculus courses before and after use of the standardized multiple-choice departmental final (SMCDF) examination at the end of the Calculus I course. Mean scores of all Calculus II-DEPT students were compared with mean scores of Calculus II-INST students completing the course. Mean scores were also compared at the Calculus III and IV levels. Overall means and standard deviations were calculated for all three subsequent levels of calculus for both the DEPT and INST groups (see Table 4.4).

The analysis in Table 4.4 indicated that the overall mean grade for all Calculus II records lowered from 2.48 for the DEPT group to 2.15 for the INST group (a decline of .33). For records where ACT-Mathematics scores were provided, the mean grade lowered from 2.50 for the DEPT group to 2.13 for the INST group (a decline of .37). For Calculus III the mean grade for all records went up from 2.48 for the DEPT group to 2.64 for the INST group (an increase of .16), for records where ACT-Mathematics scores were provided, the mean grade increased from 2.47 for the DEPT group to 2.64 for the INST group (an increase of .18). For Calculus IV the mean grade for all records lowered from 3.06 for the DEPT group to 2.89 for the INST group (a decline of .17), for records where ACT-Mathematics scores were provided, the mean grade lowered from 3.12 for the DEPT group to 2.89 for the INST group (a decline of .23). The standard deviations indicate that in Calculus IV the scores centered more on the mean score of approximately 3 for both groups while in Calculus II and III the scores showed more dispersion.

Table 4.4
Descriptive Statistics for DEPT and INST Groups

DEPT-INST		N	Mean	Std. Deviation	S. E. of Mean
Calculus II	DEPT	338(310)*	2.48(2.50)	1.278(1.259)	.070(.07)
	INST	458(428)	2.15(2.13)	1.377(1.382)	.064(.07)
Calculus III	DEPT	187(175)	2.48(2.47)	1.175(1.168)	.086(.09)
	INST	227(211)	2.64(2.65)	1.297(1.279)	.086(.09)
Calculus IV	DEPT	127(121)	3.06(3.12)	1.060(1.010)	.094(.09)
	INST	160(150)	2.89(2.89)	1.099(1.094)	.087(.09)

*Statistic of all records (statistic of records with known ACT-Mathematics scores).

A breakdown of the means and standard deviations for the DEPT and INST groups according to the ACT-Mathematics Standards for Transition ranges for Calculus II are presented in Table 4.5A (below). Tables 4.6A, and 4.7A refer to Calculus III and Calculus IV and are presented in the analysis on following pages.

Table 4.5A showed that the group means increased along the ACT-Mathematics Standards for Transition ranges in Calculus II, as would be expected. The effect size (ES), or how much the predictor variables affected the dependent variable, can be calculated using Cohen's definition as the difference between the means of the two groups divided by the standard deviation of the data set. This calculation is then converted to percentiles for interpretation. Using effect size provides a measure of outcome on a more familiar scale. A small effect size is around .20, medium is around

Table 4.5A

Descriptive Statistics for Calculus II Groups by ACT Score Ranges

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.00	1.183	11
	INST	1.15	1.292	27
	Total	1.39	1.306	38
ACTm=20-23	DEPT	1.98	1.254	55
	INST	1.38	1.400	58
	Total	1.67	1.359	113
ACTm=24-27	DEPT	2.45	1.328	116
	INST	2.03	1.301	145
	Total	2.21	1.327	261
ACTm=28-32	DEPT	2.77	1.093	119
	INST	2.49	1.278	182
	Total	2.60	1.214	301
ACTm=33-36	DEPT	3.33	1.323	9
	INST	3.38	.885	16
	Total	3.36	1.036	25
Total	DEPT	2.50	1.259	310
	INST	2.13	1.382	428
	Total	2.29	1.343	738

.50 and large is .80. For Calculus II, ES=.28 is considered a small effect size. According to Cohen (1988) this effect size in Calculus II implicated that the score of the average person in the DEPT group exceeded the scores of 62% of the INST group.

Table 4.5B

Table of Means for Calculus II Group by ACT Score Ranges

	<=19	20-23	24-27	28-32	33-36	Marginal
DEPT	2.00	1.98	2.45	2.77	3.33	2.50
INST	1.15	1.38	2.03	2.49	3.38	2.13
Overall	1.39	1.67	2.21	2.60	3.36	2.29
	-.85*	-.60	-.32	-.28	+.05	-.37

* A (-) Indicates means decreased from DEPT group to INST group.
A (+) Indicates means increased from DEPT group to INST group.

Table 4.5B displays a slightly different arrangement of the means for clarification purposes. An analysis of these means revealed that the total means for Calculus II grades consistently increased along the ACT-Mathematics Standards for Transition ranges starting at the lower level with an increase of .38, then an increase of .54, next an increase of .39, and finally .76. This, of course, would be expected since a higher student ACT-Mathematics scores would more than likely produce higher means. When comparisons were made along the same ranges, but checking for differences between groups, the INST group mean was smaller at all but the highest end of the ACT Score ranges, starting at the lower end by .85, .60, .32, and .28, but for the ACTm=33-36 range the INST group mean was .05 higher than the DEPT group mean. For a graphical representation of this data see Appendix F.

The same data were calculated for Calculus III (Table 4.6A). The group means also increased along the ACT-Mathematics Standards for Transition ranges, except for the ACTm=20-23 (ACT-Mathematics Standards for Transition range 20-23) there was a decrease compared to the previous range. The low number of records at this range could have affected the statistical analysis. When calculating the effect size for the Calculus III

Table 4.6A

Descriptive Statistics for Calculus III Groups by ACT Score Ranges

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.00	1.414	8
	INST	2.17	.983	6
	Total	2.07	1.207	14
ACTm=20-23	DEPT	2.07	1.172	30
	INST	1.80	1.508	20
	Total	1.96	1.309	50
ACTm=24-27	DEPT	2.53	1.112	62
	INST	2.55	1.248	76
	Total	2.54	1.185	138
ACTm=28-32	DEPT	2.56	1.156	71
	INST	2.86	1.184	96
	Total	2.74	1.178	167
ACTm=33-36	DEPT	3.75	.500	4
	INST	3.23	1.301	13
	Total	3.35	1.169	17
Total	DEPT	2.47	1.168	175
	INST	2.65	1.279	211
	Total	2.57	1.232	386

data, ES=.15 but in the direction of the INST group. This interpretation revealed that the score of the average person in the INST group exceeded the scores of 56% of the DEPT group. Again, according to Cohen, a very small effect size.

Table 4.6B

Table of Means for Calculus III Groups by ACT Score Ranges

	<=19	20-23	24-27	28-32	33-36	Marginal
DEPT	2.00	2.07	2.53	2.56	3.75	2.47
INST	2.17	1.80	2.55	2.86	3.23	2.65
Overall	2.07	1.96	2.54	2.74	3.35	2.57
	+.17*	-.27	+.02	+.30	-.52	+.18

\- Indicates means decreased from DEPT group to INST group.
\+ Indicates means increased from DEPT group to INST group.

The means were rearranged once more for clarification in Table 4.6B. Analysis of these means revealed that the overall means for Calculus III at the lower end first went down by .11, then increased consistently along the ACT-Mathematics ranges by .58 then .20 and finally by .61 as would be expected. When comparisons were made along the same ranges but checking for differences between groups, the INST group means exhibited no consistent pattern (+.17, -.27, +.02, +.30 and -.52).

For Calculus IV (Table 4.7A), the group means had a similar pattern. Again a small number of records at the low ranges may account for some discrepancies. For Calculus IV the effect size was calculated (ES=.22), a small effect size. Cohen’s interpretation being that the score of the average person in the DEPT group exceeded the scores of 58% of the INST group.

In this study effect size was small. Stevens (1990) mentioned that if effect size is small or medium, then a large group size is needed to detect these effects, i.e., to have adequate power. Most evaluations in social science research find small and medium effect sizes.

Table 4.7A

Descriptive Statistics for Calculus IV Groups by ACT Score Ranges

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT INST Total	2.67 2.25 2.43	.577 1.500 1.134	3 4 7
ACTm=20-23	DEPT INST Total	2.33 2.57 2.40	1.455 1.134 1.354	18 7 25
ACTm=24-27	DEPT INST Total	3.10 2.66 2.85	.889 1.066 1.014	41 56 97
ACTm=28-32	DEPT INST Total	3.35 3.10 3.20	.821 1.095 .991	55 73 128
ACTm=33-36	DEPT INST Total	4.00 3.20 3.43	.000 .789 .756	4 10 14
Total	DEPT INST Total	3.12 2.89 2.99	1.010 1.094 1.061	121 150 271

Table 4.7B

Table of Means for Calculus IV Groups by ACT Score Ranges

	<=19	20-23	24-27	28-32	33-36	Marginal
DEPT	2.67	2.33	3.10	3.35	4.00	3.12
INST	2.25	2.57	2.66	3.10	3.20	2.89
Overall	2.43	2.40	2.85	3.20	3.43	2.99
	-.42*	+.24	-.44	-.25	-.80	-.23

\- Indicates means decreased from DEPT group to INST group.

\+ Indicates means increased from DEPT group to INST group.

After looking at the rearrangement of the overall means for the Calculus IV grades in Table 4.7B, there was initially a slight lowering of .03 of the overall mean, but then followed by all increases (.45, .35 and .23) along the other ranges. When comparisons were made along the same ranges, but checking for differences between groups, the INST group means were a bit more consistent in Calculus IV (-42, +.24, -.44, -.25, and -.80 respectively).

There are a number of variables that can explain differences in student grades. Student ACT-Mathematics scores can often predict the success level in mathematics courses. Since this variable can be quantified, it can be used in the statistical procedures as an independent variable. Thus, a two-way analysis of variance (ANOVA) was run where one factor can be controlled while testing the other factor and thus compares the means of the two different non-overlapping groups DEPT and INST. The dependent variable was the course grades and the two categorical independent variables were the presence or absence of the SMCDF examination and the five levels of ACT-Mathematics Standards for Transition ranges. The analysis of variance attempts to find significant differences between group means by comparing the variances of those groups. The questions of interest to be addressed by the two-way ANOVA were:

Question #1 Was there a *main effect for* DEPT-INST? That is, did the presence of the SMCDF examinations in a first year calculus course make a significant difference in student grades in subsequent courses?

Question #2 Was there a *main effect for* ACT-scores? Did students from different levels of ACT-Mathematics Standards for Transition ranges differ significantly in their grades in subsequent calculus courses?

Question #3 Was there a DEPT-INST *by* ACT math scores *interaction?* Was the influence of the two variables idiosyncratic, such that the SMCDF examination had one effect at one level of ACT-Mathematics Standards for Transition range but a different effect at a different range?

For this study the *a priori* alpha level of 0.05 was set. Thus, tests having an alpha of 0.05 or below rejected the null hypothesis.

The analysis of variance is based on the assumptions of each score's independence of the other, homogeneity of variance, and normally distributed scores around the mean. The first assumption is clearly evident in this study. Addressing the assumption of homogeneity of variance, the total variance can be explained by (a) error of within-group variability, (b) variability due to DEPT or INST group membership, (c) variability due to ACT math scores, and (d) interaction effects between the two factors. Controlling for error variance increases the sensitivity (power) of a test; thus Levene's Test of Equality of Error of Variances was conducted (see Table 4.8).

Table 4.8

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig
Calculus II	2.010	9	728	.036
Calculus III	1.132	9	376	.339
Calculus IV	3.470	9	261	.000

The results showed a significant difference in the error of variances (p=.036) at the Calculus II level, no significant differences in the error of variances (p=.339) at the Calculus III level, yet again a significant difference (p=.000) at the Calculus IV level. Thus for Calculus II and Calculus IV data, Levene's Test indicated the variances were not homogeneous, so the assumption for the analysis of variance was not met. In order to deal with this problem, the researcher attempted to transform the data using a logarithmic transformation to the data set and again checked for homogeneity of variance. This approach did not rectify the problem. Since the data had non-homogeneous variances, the researcher needed to reduce the possibility of making a Type I error (rejecting the null hypotheses when it is true). For this reason the level of alpha was reset at .01 for Calculus II and Calculus IV.

Distributions for measures of normality were then checked. Table 4.9 shows that for all three levels of calculus the kurtosis and skewness values of the entire data set with reported ACT-Mathematics scores fell between ± 1.0. These are considered valid for most psychometric purposes. Note that the DEPT group at the Calculus IV level showed a positive value for Kurtosis indicating that the scores were centered and more clustered around a mean of 3.12. When testing for normality for the data separating DEPT group from the INST group, all data sets showed negative skewness with only the Calculus IV-DEPT data showing a larger negative value, but this was still in an acceptable range.

Table 4.9

Descriptive Statistics for Calculus II, III, and IV

	N	Mean	Std Dev.	Variance	Skewness		Kurtosis	
					Statistic	Std.Error	Statistic	Std.Error
Calculus II	738	2.29	1.343	1.803	-.352	.090	-.980	.180
Calculus III	386	2.57	1.232	1.518	-.643	.124	-.478	.248
Calculus IV	271	2.99	1.061	1.126	-.997	.148	.427	.295
Calculus II								
DEPT	310	2.50	1.259	1.584	-.549	.138	-.608	.276
INST	428	2.13	1.382	1.910	-.203	.118	-1.142	.235
Calculus III								
DEPT	175	2.47	1.168	1.365	-.481	.184	-.455	.365
INST	211	2.65	1.279	1.637	-.788	.167	-.403	.333
Calculus IV								
DEPT	121	3.12	1.010	1.020	-1.322	.220	1.544	.437
INST	150	2.89	1.094	1.197	-.783	.198	-.097	.394

This indicated that there were a larger number of grades distributed towards the higher end, meaning the A's, B's and C's. The kurtosis values were all in the negative direction except for the Calculus IV-DEPT data, which showed a large positive kurtosis value as compared to the other values. This indicated that for the Calculus IV-DEPT data the grades were more clustered around the mean instead of spread out away from the mean.

The ANOVA, which is a normal distribution-based test, was still used in this study since the sample sizes were all large enough at the Calculus I, II and III level to justify the assumption of normality. This is based on the central limit theorem which states that as sample size increases, the shape of the sampling distribution becomes normal (Howell, 1997). Tables 4.10, 4.11, and 4.12 present the results of the ANOVA tests for each level of calculus under investigation.

To answer question #1 about the main effect of DEPT-INST or the between groups variance, the results (see Table 4.10) showed there was a statistically significant

Table 4.10

ANOVA (Tests of Between-Subjects Effects) for Calculus II

Dependent Variable: Calculus II grades

Source	Type III Sum Of Squares	df	Mean Square	F	Sig	Partial Eta Squared	Observed Power
Corrected Model	166.358	9	18.484	11.573	.000	.125	1.000
Intercept	1438.914	1	1438.914	900.913	.000	.553	1.000
DEPT-INST	12.174	1	12.174	7.622	.006	.010	.787
ACT_GRP	118.277	4	29.569	18.513	.000	.092	1.000
DEPT-INST* ACT_GRP	4.962	4	1.241	.777	.540	.004	.251
Error	1162.742	728	1.597				
Total	5190.000	738					
Corrected Total	1329.100	737					

a. Computed using alpha = .01
b. R Squared = .125 (Adjusted R Squared = .114), for all explained variations.

difference (p=.006) in Calculus II grades when there was a SMCDF examination in Calculus I ignoring any effect of ACT-Mathematics scores. The observed power value of .787 indicates that there is about a 78% likelihood of finding a statistically significant difference of the same magnitude as that observed between the DEPT and INST groups in any particular sample of the same size, using the same alpha (Type I risk) level of .01.

For Calculus III (see Table 4.11) the results reported no statistically significant difference (p=.766) with only a 6% chance of finding a statistically significant difference of the same magnitude as that observed between DEPT and INST groups for any particular sample of the same size, using the same alpha (Type I risk) level.

Table 4.11

ANOVA (Tests of Between-Subjects Effects) for Calculus III

Dependent Variable: Calculus III grades

Source	Type III Sum Of Squares	df	Mean Square	F	Sig	Partial Eta Squared	Observed Power
Corrected Model	42.724	9	4.747	3.294	.001	.073	.983
Intercept	862.268	1	862.268	598.303	.614	.553	1.000
DEPT-INST	.117	1	.117	.081	.766	.000	.059
ACT_GRP	36.454	4	9.113	6.324	.000	.063	.989
DEPT-INST* ACT_GRP	4.667	4	1.167	.810	.520	.009	.259
Error	541.887	376	1.441				
Total	3134.000	386					
Corrected Total	584.611	385					

a. Computed using alpha = .05
b. R Squared = .073 (Adjusted R Squared = .051), for all explained variations.

Likewise in Calculus IV (see Table 4.12), the grades again showed no statistically significant difference with (p=.140) and a 31% chance of finding a statistically significant difference of the same magnitude as that observed between DEPT and INST groups for any particular sample of the same size, using an alpha (Type I risk) level of .01.

Table 4.12

ANOVA (Tests of Between-Subjects Effects) for Calculus IV

Dependent Variable: Calculus IV grades

Source	Type III Sum Of Squares	df	Mean Square	F	Sig	Partial Eta Squared	Observed Power
Corrected Model	30.326	9	3.370	3.214	.001	.100	.979
Intercept	708.092	1	708.092	675.336	.000	.721	1.000
DEPT-INST	2.299	1	2.299	2.193	.140	.008	.314
ACT_GRP	19.395	4	4.849	4.625	.001	.066	.945
DEPT-INST* ACT_GRP	2.715	4	.679	.647	.629	.010	.210
Error	273.659	261	1.049				
Total	2731.000	271					
Corrected Total	303.985	270					

a. Computed using alpha = .01
b. R Squared = .100 (Adjusted R Squared = .069), for all explained variations.

Question #2 in the ANOVA test considered the main effect of ACT-GRP or the within group variances. The results (see Tables 4.11, Table 4.12, and Table 4.13) showed that there was a statistically significant difference (p=.000, p=.000 and p=.001) amongst the five levels of Standards for Transition in ACT-Mathematics scores for all reported scores from both the DEPT and INST groups combined. The observed power was 100%, 99% and 95% respectively of the chance of finding a significant difference of the same magnitude as that observed in grades at the five levels for this sample size, using the same alpha (Type I risk) level. These results just verify that a student's ACT-

Mathematics score tends to correlate with his or her level of success, which is the intended purpose of ACT tests.

For question #3 the DEPT-INST by ACT-Mathematics ranges interaction, the *p*-values were p=.540 (see Table 4.10), p=.520 (see Table 4.11), and p=.629 (see Table 4.12) indicating no significant influence between the two variables. The F statistics are .777, .810 and .647 respectively for the interaction effect. This is not to say that there was no interaction at all between the different ACT-Mathematics ranges and the two groups, but that it was not statistically significant. Thus, for all five ACT-Mathematics score ranges, there was no change in the effects on student grades that were already found for the two groups DEPT and INST at all three levels of Calculus.

After close examination of the data set going back to Figure 4.1, notice must be made of the large number of audits at the Calculus I level for the DEPT group compared to the INST group. This was largely due to a change of policy at the university in the fall semester of 1995 when a super drop policy was converted to a super audit policy (MSU Registrar's Office Tool Kit, 1995). This policy was totally eliminated in the fall of 1997. Table 4.13 shows the numbers and percents of students auditing a calculus course. The super audit policy was in effect the same time period from which the DEPT group data was obtained. These calculations show there was a much higher percent of audits in the DEPT group at the first three levels of calculus. The largest differences were for Calculus I and Calculus II where the DEPT groups had 19.63% and19.08% rates of audits versus the INST groups that had only .15% and 0.20% rate of audits.

After further investigation, it was also observed that the INST group had a large number of F's and W's when compared to the DEPT group. When F grades were

Table 4.13

Percent of Students Auditing Calculus

	Group	Audit	Total	Percent	Difference
Calculus I	DEPT	193	983 (892)*	19.63% (21.64%)	
	INST	2	1315 (1016)	00.15% (00.20%)	19.48% (21.44%)
Calculus II	DEPT	83	435 (396)	19.08% (20.96%)	
	INST	1	499 (414)	00.20% (00.24%)	18.88% (20.72%)
Calculus III	DEPT	28	223 (208)	12.56% (13.46%)	
	INST	0	241 (216)	0.00% (00.00%)	12.56% (13.46%)
Calculus IV	DEPT	1	132 (127)	0.76% (00.79%)	
	INST	0	165 (159)	0.00% (00.00%)	0.76% (00.79%)

*Total = Number of records less S records or DNE (non existent) records.
(Percent of Students Auditing Calculus when F grades are excluded)

excluded, these differences in percents remained consistent. From Figure 4.1, it can be speculated that the super audit policy had a major effect on the fewer number of F grades in the DEPT group, particularly in Calculus I and II. Table 4.14 shows how the total

Table 4.14

Frequency and Percent of F, AU, and W Grades in the DEPT and INST Groups

	F	AU	W	Total	Percent of Group
Calculus I					
DEPT	91(29.0%)	193(62.0%)	28(09.0%)	312(100%)	31.7%
INST	299(67.0%)	2(00.4%)	145(32.6%)	446(100%)	33.9%
Calculus II					
DEPT	39(29.0%)	83(61.0%)	14(10.0%)	136(100%)	31.3%
INST	85(67.5%)	1(00.8%)	40(32.7%)	126(100%)	25.3%
Calculus III					
DEPT	15(29.4%)	28(55.0%)	8(15.6%)	51(100%)	22.9%
INST	25(64.0%)	0(00.0%)	14(36.0%)	39(100%)	16.2%
Calculus IV					
DEPT	5(50.0%)	1(10%)	4(40%)	10(100%)	7.6%
INST	6(55%)	0(0%)	5(45%)	11(100%)	6.7%

number of F's, AU's and W's were combined and percents were calculated for comparison. The DEPT group and the INST group showed these grades to account for 31.7% and 33.9% respectively of the total grades excluding the S (Satisfactory) grades. This shift from AU's to F's is attributable to the abolishment in the fall of 1997 of the super audit policy that was put into effect in August of 1995 (MSU Registrar's Office, personal communication, March 5, 2003).

In the two-way ANOVA, F grades were included in the analysis. Since there was such an obvious shift in the number of these F grades due to the super audit policy, the validity of the ANOVA statistics was questioned. In order to increase the validity of the study, the statistical procedures were repeated after excluding F grades, along with AU and W grades, from the data. This would possibly eliminate some of the effect on the results caused solely from the "super audit" policy.

Levene's Test of Equality of Variance was performed once again (see Table 4.15), but no changes were observed. For Calculus II and IV, the variances differed significantly p=.012 and p=.005 respectively. Thus for Calculus II and Calculus IV data,

Table 4.15

Levene's Test of Equality of Error Variances (Excluding F Grades)

	F	df1	df2	Sig
Calculus II	2.371	9	614	.012
Calculus III	1.055	9	340	.396
Calculus IV	2.698	9	252	.005

Levene's Test once again indicated non-homogeneous of variances violating the assumption for the ANOVA. Transformation of the data using a logarithmic transformation was attempted, but this approach once again did not resolve the problem. In order to reduce the possibility of making a Type I error (rejecting the null hypotheses when it is true), the level of alpha was set at .01 for Calculus II and Calculus IV.

Descriptive statistics were run on the reduced data set (Table 4.16). The DEPT group mean for Calculus II still showed to be higher than the INST group mean, but the difference was not as great. There was a difference of .37 between the means when

Table 4.16

Descriptive Statistics for Calculus II, III, and IV (Excluding F Grades)

	N	Mean	Std. Dev.	Variance	Skewness		Kurtosis	
					Statistic	Std.Error	Statistic	Std.Error
Calculus II								
DEPT	277	2.80	.968	.937	-.237	.146	-.999	.292
INST	347	2.63	1.021	1.043	-.064	.131	-1.142	.261
Calculus III								
DEPT	161	2.68	.951	.905	-.119	.191	-.938	.380
INST	189	2.96	.953	.908	-.596	.177	-.576	.352
Calculus IV								
DEPT	117	3.22	.842	.709	-.971	.224	.428	.444
INST	145	2.99	.968	.937	-.545	.201	-.781	.400

F grades were included, but excluding the F grades the difference showed to be .17. For Calculus III, the gap was also narrowed between the two groups, with the INST group showing once again to have the higher mean. There was a difference of .18 when the F grades were included, and .28 when the F grades were excluded. For Calculus IV, even though the means were slightly higher, the gap between the means for the new statistics did not change.

The results of the ANOVA tests on the data that excluded F grades are presented in Tables 4.17, 4.18, and 4.19. Furthermore, a complete set of descriptive statistics, two-way ANOVAs, and interaction line plots for Calculus II, III and IV when F grades were excluded from the data set can be found in Appendix H.

Hypothesis #1.

The first null hypothesis in the study stated as follows:

HO_1 : There will be no statistically significant difference in the mean Calculus II course grades between Calculus II students who have taken a Calculus I SMCDF examination (DEPT-II) and Calculus II students who have not taken a Calculus I SMCDF examination (INST-II).

In Calculus II (see Table 4.17), with an alpha level of .01, there was no statistically significant difference in the main effect of DEPT-INST or variance between the DEPT group and INST group (*p*=.044) with an observed power of .521 when F grades were excluded. The main effect of ACT-GRP or variances between the different ACT ranges was statistically significant (*p*=.000) as would be expected, and there was no significant DEPT-INST by ACT-Mathematics ranges interaction effect (*p*=.889).

Based on the statistical information of the initial ANOVA test before F grades were excluded, there was a statistically significant difference (*p*=.006) in the mean grades with a 79% chance of not making a Type II error (failing to reject the null hypothesis when it is in fact false). When the F grades were excluded the ANOVA test did not show a statistically significant difference (*p*=.044) in the mean grades with a 52% chance of not making a Type II error. These results were considered to be inconclusive and thus a statistically significant difference was not found between the DEPT and the INST group means for Calculus II with alpha set at .01. As a result, this null hypothesis was not rejected (Howell, 1997). One can conclude from this that students at MSU who took a standardized multiple-choice departmental final examination at the end of their Calculus I course did not make higher grades in Calculus II than those students who were not required to take this type of examination at the end of the course.

Table 4.17

Two Way ANOVA for Calculus II (excluding F grades)

Dependent Variable: Calculus II grades

Source	Type III Sum Of Squares	df	Mean Square	F	Sig	Partial Eta Squared	Observed Power
Corrected Model	36.534	9	4.059	4.245	.000	.059	.998
Intercept	1669.587	1	1669.587	1745.761	.000	.740	1.000
DEPT-INST	3.888	1	3.888	4.065	.044	.007	.521
ACT_GRP	29.087	4	7.272	7.603	.000	.047	.997
DEPT-INST* ACT_GRP	1.081	4	.270	.283	.889	.002	.113
Error	587.209	614	.956				
Total	5190.000	624					
Corrected Total	623.744	623					

a. Computed using alpha = .01
b. R Squared = .059 (Adjusted R Squared = .045), for all explained variations.

Hypothesis #2.

The second null hypothesis of the study stated:

HO_2 : There will be no statistically significant difference in the mean Calculus III course grades between Calculus III students who have taken a Calculus I SMCDF examination (DEPT-III) and Calculus III students who have not taken a Calculus I SMCDF examination (INST-III).

For Calculus III (Table 4.18), no statistically significant difference was found in the main effect DEPT-INST or between groups variance (*p*=.680), and no significant DEPT-INST by ACT-Mathematics ranges interaction effect was found (*p*=.871) when the

F grades were excluded. The observed power value of .070 suggested there was only a 7% chance of finding a significant difference of the same magnitude as that observed between DEPT and INST groups in any particular sample of the same size, using the same alpha (Type I risk) level.

Table 4.18

Two Way ANOVA for Calculus III (excluding F grades)

Dependent Variable: Calculus III grades

Source	Type III Sum Of Squares	df	Mean Square	F	Sig	Partial Eta Squared	Observed Power
Corrected Model	29.850	9	3.317	3.855	.001	.093	.994
Intercept	969.318	1	969.318	1126.582	.000	.768	1.000
DEPT-INST	.146	1	.146	.170	.680	.000	.070
ACT_GRP	21.536	4	5.384	6.257	.000	.069	.988
DEPT-INST* ACT_GRP	1.069	4	.267	.311	.871	.004	.120
Error	292.538	340	.860				
Total	3134.000	350					
Corrected Total	322.389	349					

a. Computed using alpha = .05
b. R Squared = .093 (Adjusted R Squared = .069), for all explained variations.

Based on the statistical results of the initial ANOVA test (p=.766 and a 6% chance of not making a Type II error) and the ANOVA test when F grades were excluded (p=.680 and a 7% chance of not making a Type II error), no significant difference was found between the DEPT and the INST group means at the Calculus III level. This second null hypothesis was therefore retained. Students at MSU who took a standardized multiple-choice departmental final examination at the end of their Calculus I course did not perform higher in Calculus III than those students who were not required to take this type of examination at the end of Calculus I.

Hypothesis #3.

The third null hypothesis of the study stated:

HO_3 : There will be no statistically significant difference in the mean Calculus IV course grades between Calculus IV students who have taken a Calculus I SMCDF examination (DEPT-IV) and Calculus IV students who have not taken a Calculus I SMCDF examination (INST-IV).

For Calculus IV (Table 4.19) the ANOVA showed some different results when the F grades were excluded from the analysis. Since the data set showed non-homogeneous variances, the alpha level was set at .01. At this level, the main effect of DEPT and INST or between groups variance did not show a statistically significant difference (*p*=.046).

Table 4.19

Two Way ANOVA for Calculus IV (excluding F grades)

Dependent Variable: Calculus IV grades

Source	Type III Sum Of Squares	df	Mean Square	F	Sig	Partial Eta Squared	Observed Power
Corrected Model	22.624	9	2.514	3.199	.001	.103	.979
Intercept	737.630	1	737.630	938.845	.000	.788	1.000
DEPT-INST	3.160	1	3.160	4.022	.046	.016	.515
ACT_GRP	17.433	4	4.358	5.547	.000	.081	.976
DEPT-INST* ACT_GRP	1.207	4	.302	.384	.820	.006	.138
Error	197.991	252	.786				
Total	2731.000	262					
Corrected Total	220.615	261					

a. Computed using alpha = .01
b. R Squared = .103 (Adjusted R Squared = .070), for all explained variations.

Again, there was no significant DEPT-INST by ACT-Mathematics ranges interaction effect (p=.820) when F grades were excluded. The observed power value of the main effect was .515, suggesting there was about a 52% chance of finding a statistically significant difference of the same magnitude as that observed between DEPT and INST groups in any particular sample of the same size, using the same alpha (Type I risk) level of .01.

Based on the statistical results found in the ANOVA which excluded F grades to give the statistics more validity, a statistically significant difference (p=.046 using an alpha level of .01) was not found between the DEPT and INST group means at the Calculus IV level. Consequently, this null hypothesis was not rejected. Students at MSU who took a standardized multiple-choice departmental final examination at the end of their Calculus I course did not make higher grades in Calculus IV than those students who were not required to take this type of examination at the end of Calculus I.

Hypotheses #4 through #8

The research question in this study also stated that if a difference was found, is this difference consistent along varying levels of students' ACT-Mathematics assessment scores? No statistically significant differences were found between the two groups for any level of calculus, but the question can still be asked for the different levels of ACT Mathematics scores Standards for Transition ranges. The original ANOVA showed no significant interaction effect of DEPT-INST by ACT-Mathematics score ranges (p=.540, p=.520, and p=.629 for Calculus II-IV respectively) nor a significant interaction effect when F grades were excluded (p=.889, p=.871, and p=.820). A summary table of sample sizes was constructed for the ACT Mathematics Standards for Transition ranges (Table 4.20).

Table 4.20

N-values by ACT-Mathematics Standards for Transition Score Ranges

ACT Groupings	DEPT-INST	N-value Original ANOVA			N-value ANOVA excluding F's		
		II	III	IV	II	III	IV
ACTm<=19	DEPT	11	8	3	9	7	3
	INST	27	6	4	15	6	4
		II	III	IV	II	III	IV
ACTm=20-23	DEPT	55	30	18	43	27	15
	INST	58	20	7	34	14	7
		II	III	IV	II	III	IV
ACTm=24-27	DEPT	116	62	41	102	59	41
	INST	145	76	56	117	68	54
		II	III	IV	II	III	IV
ACTm=28-32	DEPT	119	71	55	115	64	54
	INST	182	96	73	165	89	70
		II	III	IV	II	III	IV
ACTm=33-36	DEPT	9	4	4	8	4	4
	INST	16	13	10	16	12	10

The fourth null hypothesis stated:

HO_4 : There will be no statistically significant differences in the mean Calculus II, III, and IV course grades of students having an ACT-Mathematics assessment score of 16-19 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

Based on an alpha level of .05 (see Table 4.21), Cohen's definition for a medium effect size, and a selected power of .80 (20% chance of making a Type II error), the sample size needed, or n-value, was calculated to be about 31 (Howell, 1997).

Table 4.21

Estimated Required Sample Sizes

	α = .01	α = .05
power = .80	δ = 3.40	δ = 2.80
Small Effect Size d=.20	N = 289	N = 196
Medium Effect Size d=.50	N = 46	N = 31

Calculating $N = (d / d)^2$

For the ACTm<19 range, each of the comparison groups (DEPT-II and INST-II through DEPT-IV and INST-IV) had an n-value ranging from only 3 to 27 student scores (Table 4.20). For this reason the validity of the results was questioned at this ACT range. Consequently, the fourth null hypothesis was not rejected. Thus for students with ACT-Mathematics scores below 19, it was not possible to determine if a standardized multiple-choice departmental examination at the end of a Calculus I course made any difference in their Calculus II, III or IV grades.

The fifth null hypothesis stated:

HO_5: There will be no statistically significant differences in the mean for Calculus II, III, and IV course grades of students having an ACT-Mathematics assessment score of 20-23 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

For Calculus III the data set had homogeneous variances and the alpha level was set at .05. The ANOVA tests found no statistically significant differences in the mean grades between the groups. For a medium effect size, the sample sizes of the groups needed to be at least 31, but for this ACT-Mathematics range the sample sizes of the groups were too small for any statistically significant differences to be found. For Calculus II and IV where the alpha level had to be set at .01, for a medium effect size the sample size needed to be at least 46 and for a small effect size the sample size needed to be at least 289 (see Table 4.21). Referring to Table 4.20, the n-values for the Calculus II comparison groups were 55 and 58 for the original ANOVA, sufficient for a medium effect size at an alpha level of .01, but not for the ANOVA excluding F's where the sample sizes were too small. For the Calculus IV, the sample sizes were all too small for either effect size. The conclusion was made that for students who had ACT-Mathematics scores range from 20-23, the sample sizes were too small to find a significant difference at the Calculus II, III, and IV levels between the DEPT and INST groups. Consequently, the fifth null hypothesis was not rejected for Calculus II, III and IV. In simple terms, students with ACTm=20-23 for this sample were not more successful in Calculus II, III, or IV when they took a standardized multiple-choice departmental final examination at the end of their Calculus I course.

The sixth null hypothesis stated:

HO_6 : There will be no statistically significant differences in the mean

Calculus II, III, and IV course grades of students having an ACT-Mathematics assessment score of 24-27 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

For Calculus III, the ANOVA tests found no statistically significant differences in the mean grades between the groups. The sample sizes were all found to be over 31 which was needed at the alpha level of .05, thus for this ACT-Mathematics range no statistically significance was found. For Calculus II and IV where the alpha level had to the set at .01, a sample size of 46 was needed for a medium effect size and a sample size of 289 was needed for a small effect size. From Table 4.20 the n-values at this ACT-Mathematics range were all below 289, but most were above 46. The sample sizes were large enough for a medium effect size, but not for a small effect size. Thus, for students in the ACT-Mathematics score range of 24-27, sample sizes were too small to determine a statistically significant difference between the DEPT and INST groups all three levels of calculus. Consequently, this null hypothesis was not rejected. One can conclude that these students with ACTm=24-27 who took a standardized multiple-choice departmental final examination at the end of Calculus I were not more successful in Calculus II, III and IV.

The seventh null hypothesis stated:

HO_7 : There will be no statistically significant differences in the mean Calculus II, III, and IV course grades of students having an ACT-Mathematics assessment score of 28-32 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

For Calculus III, the sample sizes at this ACT-Mathematics range were all adequate. The ANOVA tests found no statistically significant differences using the alpha level at .05. For Calculus II and IV where the alpha level had to be set at .01, the sample sizes (see Table 4.20) were large enough for a medium effect size (N=46), but not for a small effect size (N=289). The groups of students who had ACT-Mathematics scores ranging from 28-32, had sample sizes too small to find a statistically significant difference, consequently this null hypothesis was rejected. Once again, one can conclude that students with ACTm=28-32, who took a standardized multiple-choice departmental final examination at the end of Calculus I were no more successful in Calculus II, III or IV.

The eighth null hypothesis stated:

HO_8 : There will be no statistically significant differences in the mean Calculus II, III, and IV course grades of students having an ACT-Mathematics assessment score of 33-36 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

Referring to Table 4.20, at this ACT-Mathematics score range of 33-36 the n-values were all quite small. For students whose ACT-Mathematics scores ranged between 33 and 36, no significant difference between the DEPT and the INST groups was determined with such small sample sizes. Consequently, this null hypothesis was not rejected for the two comparison groups. For students with ACTm=33-36, it was not possible to determine if a standardized multiple-choice departmental final examination at the end of a Calculus I course made any difference in their Calculus II, III or IV grades.

Summary

The purpose of this study was to compare the student grades of those who took a departmental multiple-choice final examination and those who did not take the departmental examination but instead were given an instructor generated examination. Comparisons were further made along five different ACT-Mathematics Standards for Transition ranges. Levene's test was used to check for homogeneity of variances. For Calculus II and Calculus IV the data sets were found to be non-homogeneous. A two-way ANOVA was used to compare the means at an alpha level of .05 for Calculus III but the alpha level was set at .01 for Calculus II and IV to reduce the Type I error (rejecting the null hypotheses when it is true).

When the data were further examined, it was found that the DEPT group had a high number of AU or audit records and the INST group had a high number of F records. A "super audit" policy was in effect during the time period from which DEPT records were obtained which could have caused the internal validity of this study to be weakened. To improve the validity of the study, the ANOVA tests were repeated excluding the F grades to balance the high number of audits during the DEPT group time frame.

Originally, a significant difference was found between the means of the Calculus II-DEPT and the Calculus II-INST groups. When the F records were excluded, the results changed and showed no statistically significant difference.

Table 4.22 summarizes the findings of this study. Even though the two-way ANOVA did not show any interaction effect between the ACT-Mathematics scores and the two groups at all three levels of calculus, the sample size at each ACT-Mathematics Standards for Transition range had to be considered.

Table 4.22

Summary of Statistically Significant Differences Found

ACT Groupings	Original ANOVA			ANOVA excluding F's		
ACTm<=19	II	III	IV	II	III	IV
ACTm=20-23						
ACTm=24-27	Yes	No	No	No	No	No
ACTm=28-32						
ACTm=33-36						

For this study it can be summarized; for Calculus II students who took the SMCDF examination at the end of Calculus I, there was a statistically significant difference found in their level of success in the course, when F grades were included in the data set. This significance was not found at the Calculus III and IV levels. When F grades were excluded from the data set due to the effect of an audit policy, no statistically significant differences were found for any of the three levels of Calculus. Further research would be recommended using a larger number of records and eliminating the two extreme ACT-Mathematics Standards for Transition ranges.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter first provides a brief summary of the study under investigation. The next section discusses the findings of the study, and then is followed by some discussion and recommendations for further research based on the results in this study.

Summary

A large number of students are required to study calculus as an integral part of their major area of study. The beginning level of calculus, often referred to as Calculus I, is the foundation on which the subsequent calculus courses rely. Adequate preparation at this beginning level can provide greater opportunity for success at the subsequent levels of the course. This study examined a total of 2560 (1063 in DEPT and 1497 in INST) Calculus I student records spanning a period of 8 semesters at a major university and followed them through the four levels of calculus required for their majors. From the initial Calculus I groups, 256 (114 in DEPT and 142 in INST) students or 10% (10.7% in DEPT and 9.5% in INST) completed and passed the final course in the four-course sequence. These percentages changed to 11.6% for the DEPT group and 10.8% for the INST group when S grades were excluded from the study. These percents do not indicate that all students not finishing the sequence failed the courses, some were not required by their major area of study to take all four of the courses in the sequence.

When multiple sections of the beginning level are offered at learning institutions and varying instructors are assigned the task of teaching the course, questions concerning the consistency in depth of understanding and material covered could arise. At one time, the Department of Mathematics and Statistics at Mississippi State University implemented a standardized multiple-choice departmental final (SMCDF) examination at the end of the Calculus I course that would address some of the inconsistencies and to assist in determining if a student was adequately prepared to go on to the next level of calculus. Along with this departmental examination, there was a policy in place that this examination was to count as 50% of the final course grade in the event the examination grade was equal to or higher than the student's class average prior to the exam and counted as 33 1/3% of the grade in the event it was lower. In the fall semester of 1997, a committee in the mathematics and statistics department made the decision to abolish the examination and instead allowed each instructor to generate their own final examination and grading policy. The purpose of this study was to determine if there was a difference in students' success in subsequent calculus courses before and after use of the standardized multiple-choice departmental final (SMCDF) examination for Calculus I.

This study utilized a causal-comparative design. This design allows researchers to examine cause and effect relationships under conditions where experimental manipulation is difficult or impossible. Many variables were involved that could have an effect on the results of this study. Some of the most obvious being the ACT-Mathematics scores of the students involved in the study, the university policy for auditing a course, teaching styles of the different instructors and professors, the changes in instructors in the department, the change in use of the graphing calculator or computer algebra systems, the changes in teaching staff in the department, and teacher grading policies, to name just a few.

It was possible to effectively control the variable of student ACT-Mathematics scores, since they were available from the university records. These scores were categorized according to Standards for Transition ranges and used as a categorical independent variable. For this reason, a two-way ANOVA statistical test was run using the ACT-Mathematics score and the presence or absence of the SMCDF examination as independent variables. The two-way ANOVA allowed the researcher to control for one of the variables while checking for differences in the other variable. After reexamination of the data set, it was found that a super audit university policy was in effect during the time period from which the DEPT scores were obtained affecting the validity of the ANOVA results. For Calculus I, the DEPT group showed a 19.63% audit rate versus a 0.5% for the INST group. The DEPT group in Calculus II showed a 24.37% rate versus 0.4% for the INST group, and the Calculus III group an 18.83% rate compared to no audits for the INST group. In Calculus IV this difference was not so extreme, with a 3.79% rate for the DEPT group and none for the other. These percentages indicated that the policy change could have had a considerable impact on the results of this study. This is because students from the DEPT group who chose to use their super audit might have otherwise had to stay in the course and receive a grade that would have been included in the analysis of the study. For this reason the F grades were eliminated from the study along with the AU records to provide more validity to the two-way ANOVA statistics.

When testing for the assumptions of the ANOVA, significant differences were found in the variances of the Calculus II and IV data indicating non-homogeneous variances. In order to reduce the Type I error and increase the power of the test, a .01 alpha level was set for the Calculus II and Calculus IV statistics. Statistically significant differences were found between the two groups at the Calculus II level (p=.006) before F grades were excluded but none were found (p=.044) when F grades were excluded. For Calculus IV (p=.140, and p=.046 respectively) no statistically significant differences were found in the mean grades for those students who took the SMCDF examination as compared to those students who took the individual instructor generated final examination. For Calculus III, the data showed homogeneous variances, thus the alpha level was set at .05. The ANOVA showed no statistically significant differences at this level. The interaction of the ACT-Mathematics scores was not found to be significant (p=.889 for Calculus II, p=.871 for Calculus III, and p= .820 for Calculus IV). Upon closer examination of the data along each of the ACT-Mathematics Standards for Transition ranges, the researcher found the sample sizes at the lower range (ACTm<19) and upper range (ACTm33-36) were too small to provide any validity to the statistics at any of the calculus levels. The sample sizes were also too small for all other ranges for Calculus II and Calculus III when the alpha level was set for .01 to reduce Type I error. For this reason, any significant difference for these ranges could not be justified.

Conclusion

The purpose of this study was to determine if there was a difference in students' success in subsequent Calculus courses before and after use of the standardized multiple-choice departmental final (SMCDF) examination for Calculus I. The null hypotheses for the study were as follows:

HO_1: There will be no statistically significant difference in the mean Calculus II course grades between Calculus II students who have taken a Calculus I SMCDF examination (Calculus II-DEPT) and Calculus II students who have not taken a Calculus I SMCDF examination (Calculus II-INST).

The difference in the mean grades of the group of Calculus II students who took the SMCDF examination in Calculus I was statistically significant from those who did not take the examination when F grades were not excluded, however when F grades were excluded the differences were no longer statistically significant. Therefore this null hypothesis was retained.

HO_2: There will be no statistically significant difference in the mean Calculus III course grades between Calculus III students who have taken a Calculus I SMCDF examination (Calculus III-DEPT) and Calculus III students who have not taken a Calculus I SMCDF examination (Calculus II-INST).

This null hypothesis was not rejected. The difference in the mean grades of the group of Calculus III students who took the SMCDF examination in Calculus I was not statistically significantly from whose who did not take the examination, whether F grades were or were not excluded.

HO_3: There will be no statistically significant difference in the mean Calculus IV course grades between Calculus IV students who have taken a Calculus I SMCDF examination (Calculus IV-DEPT) and Calculus IV students who have not taken a Calculus I SMCDF examination (Calculus IV-INST).

With the alpha level at .01, the difference in the mean grades of the group of Calculus IV students who took the SMCDF examination in Calculus I was not statistically significant from those who did not have to take such examination, whether F grades were or were not excluded. This hypothesis was also retained.

HO_4: There will be no statistically significant differences in the mean Calculus II, III, and IV course grades of students having an ACT mathematics assessment score of 16-19 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

This hypothesis was not rejected. It was not possible to determine any significant differences in mean grades between the two groups of Calculus II, III, and IV students having an ACT mathematics assessment score of 16-19 who took the SMCDF examination in Calculus I, since the sample sizes were too small.

HO_5: There will be no statistically significant differences in the mean Calculus II, III, and IV course grades of students having an ACT mathematics assessment score of 20-23 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

There was a statistically significant difference in the mean grade of the Calculus II students having an ACT Mathematics score of 20-23 who took the SMCDF examination in Calculus I before the F grades were excluded using an alpha level of .01. When F grades were excluded that difference was no longer significant. For Calculus III and IV there was no significant difference at this ACT Mathematics range. This null hypothesis was not rejected

HO_6: There will be no statistically significant differences in the mean Calculus II, III, and IV grades of students having an ACT mathematics assessment score of 24-27 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

There was a statistically significant difference in the mean grade of the group of Calculus II students having an ACT Mathematics score of 24-27 who took the SMCDF examination in Calculus I before the F grades were excluded using an alpha level of .01. When F grades were excluded that difference was no longer significant. For Calculus III and IV there was no significant difference at this ACT Mathematics range. This null hypothesis was not rejected.

HO_7: There will be no statistically significant differences in the mean Calculus II, III, and IV grades of students having an ACT mathematics assessment score of 28-32 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

There was a statistically significant difference in the mean grade of the group of Calculus II students having an ACT Mathematics score of 28-32 who took the SMCDF examination in Calculus I before the F grades were excluded using an alpha level of .01. When F grades were excluded that difference was no longer significant. For Calculus III and IV there was no significant difference at this ACT Mathematics range. This null hypothesis was not rejected.

HO_8: There will be no statistically significant differences in the mean Calculus II, III, and IV grades of students having an ACT mathematics assessment score of 33-36 between Calculus II-DEPT and Calculus II-INST, Calculus III-DEPT and Calculus III-INST, and Calculus IV-DEPT and Calculus IV-INST groups.

This hypothesis was not rejected. It was not possible to determine any significant differences in the mean of the student grades between the two groups of Calculus II, III, and IV students having an ACT mathematics assessment score of 33-36 who took the SMCDF examination in Calculus I, since the sample sizes were too small.

Discussion

Since there were no other studies found in the research literature that addressed the same research question under investigation, this study was viewed as a ground-level study. The purpose was to discover whether or not a statistically significant difference existed between the DEPT group of students who were required to take the SMCDF examination and the INST group of students who were not required to take the test. Since differences were found between the groups in the data set at the Calculus II level before F grades were excluded, but none when the F grades were removed to obtain more internal validity, recommendations would be made for further research.

Conflicting findings seem to appear in particular for the Calculus III level. First the DEPT group grade distributions were more abundant in the B's and C's, where the INST group grade distributions concentrated heavier in the A's and B's. This might have been attributed to the existence of the audit policy for the DEPT group where students might not have applied themselves as much knowing the audit policy gave them an opportunity to drop the course during the semester. Next, there was a switching back and forth between increasing and decreasing in the means along the different ACT-Mathematics Standards for Transition ranges. This might have been due to a lower sample size at those extreme ranges, where the data might not have been normally distributed. This was also seen in the line plots showing any possible interaction effect that these ACT ranges had with the two groups. These inconsistencies could have also been attributed to the content of the Calculus III course itself. Students and teachers have often remarked that the topics covered in Calculus III seem to be varied, unique and often more difficult. Topics at this level of calculus were often completely new to the student with less of a building block feature. It would have been an interesting research question to ask the students who completed all four calculus courses on their differing opinions about the concepts they learned at each level.

The original data set for this study used grades from multi-section calculus courses. A number of teachers taught these courses (see Table 3.1). Varying teaching methods, experience in teaching, competence level of the material and personality differences can influence the results of the scores in the data sets. This factor was not addressed nor controlled for in this study. Further research could be done by investigating teaching styles through the use of questionnaires, interviews, and/or observations.

A simple statistic that was calculated and found to be interesting was the rate of students completing the entire calculus sequence. From the DEPT group of 983 students receiving a Calculus I grade, 114 students passed Calculus IV. This calculates to 11.6% of the students who took Calculus I continued the sequence and completed and passed Calculus IV. From the INST group the statistic was not much different. A total of 1315 students received a Calculus I grade in this group, of which 142 students passed Calculus IV. This indicated that 10.8% of the students who took Calculus I continued the sequence and completed and passed Calculus IV. It must be noted, however, that some students who took Calculus I or Calculus II were not required to take the higher level courses of calculus for their major area of study. This information was not available from the student records; therefore, the percentages were not necessarily an indication of just failing grades along the sequence of the courses. Adjustments in these percents would have to be made, from which further research could be conducted. The question could be raised whether the corrected percentages would be acceptable failure rates.

The statistics from this study came from one institution. Do other institutions of higher learning keep data of a similar kind? Were these percentages the norm amongst other institutions? Referring to Table 4.3 on page 43 of this study, close to half of the students taking Calculus I failed the course (DEPT is 42%, INST is 46.24%). In Calculus II, over one-third of the students failed the course (DEPT is 37.93%, INST is 35.87%). In Calculus III the failure rate diminished (DEPT is 31.39%, INST is 24.07%), but still ranged about one in every three or four students failing. In Calculus IV it was less than 15% (DEPT is 13.64%, INST is 13.94%). Whether a student took a SMCDF examination in Calculus I or not, these percentages could indicate an area of concern for any educational organization.

It is important to note that significant differences were found for Calculus II, but whether these differences lasted is inconclusive. Even though these conflicting results did not clearly indicate whether a SMCDF examination made a difference in the levels of success in subsequent calculus courses, indications seem to be that at least at the Calculus II level, which is the first course immediately following the SMCDF examination, there might have been an effect. The multiple-choice format of the SMCDF examination was used to assess a large number of students. This format of testing is very time efficient, easily administered, and is the most cost-effective format of assessment possible, yet these types of tests do not evaluate reasoning skills very well. Madaus and Kellaghan (1993) and Herman (1992) listed a number of research references implicating that the multiple-choice, norm-referenced standardized tests often measure low-level knowledge and thinking skills. Teachers feeling the pressure of accountability often will teach to the test. As a result, education is being driven in the wrong direction. Performance assessment could instead be used to evaluate higher-order thinking skills. What effect would this have on the performance of the students? If there were an effect, would this be an effect that could be an influence in subsequent courses?

Some form of assessment is generally in existence in most courses. This is especially the case in mathematics courses. It has been reported that improved teacher classroom assessment practices have had consistent and sizable gains in standardized test scores (Stiggins, 1999a), yet assessment may or may not be a topic covered in a curriculum or staff development of a school or department. Furthermore, hiring standards or staff evaluations are seldom based on the assessment capabilities of an instructor (Stiggins, 1995, 1999b). Demonstrating competence in assessment as a condition to being licensed to teach is rarely a requirement. As a result, assessment or testing by the classroom teacher is very inconsistent from teacher to teacher.

Recommendations

Based on the findings and conclusions of this study, the following recommendations can be made for future studies.

Incorporating teaching styles into the study would give more validity to the statistics. Through the use of questionnaires, observations and interviews, instructors with similar teaching methods could be used to investigate the same research questions in this study. A much larger data set would be necessary to answer similar questions. Further investigation can be done by exploring how different teaching styles may affect the results in a SMCDF examination. These may even be compared to results of the same teaching styles without the use of the SMCDF examination and instead use the teacher generated examination. In either case, a much larger data set would be needed to be able to calculate any statistical test procedures that may apply to the research question of interest.

This study occurred at only one institution. Randomly selecting several other institutions with similar calculus programs to see if the data reveal the same results could direct further research into different directions. Student surveys, especially at the Calculus III levels where the descriptive and inferential statistics of this study showed conflicting results, may bring further understanding to the results of this study. Investigating if there would be any differences between students coming from different parts of the state, and methods of assessing a large number of students can also be explored.

The format of the final examination was a multiple-choice method of assessment. Further studies could investigate the possibility of consistency and accountability using a final examination that had a mixed format using multiple-choice questions and free response questions similar to the Advanced Placement examination. Research is needed to determine whether such a test at the Calculus I and possibly Calculus II levels would make a difference in the success of the students as they progressed further. Would a different format of testing bring about better understanding of the concepts and in turn produce a higher percentage of passing grades at the end of each subsequent course? These questions can be researched by various forms of longitudinal studies. From the results of this study, using students in the middle ranges of ACT Mathematics scores as subjects, and following their success level after using different formats of final examination testing could provide valuable information. It is the belief of this researcher that some form of a departmental final examination could benefit the student and the department.

In this computer age, record-keeping ability for any department should not be a difficult task. Much research can be done when records are kept current within a department. Statistical analysis can be performed instantly on many research questions that may be raised. Much can be learned about learning conditions for students, assessment methods used, effective teaching, and much more. With the data at hand, many valuable departmental decisions can easily be made.

REFERENCES

Ahrens, S. (1980). *Analysis and Classification of Entering Freshmen Mathematic Students Using Multiple Discriminate Function Analysis.* Paper presented at the Annual Meeting of the Southern Association of Institutional Research, Louisville, KY.

ACT: News: Facts about the ACT assessment. (n.d.). Retrieved February 7, 2003, from http://www.act.org/news/aapfacts.html

ACT: Research: Information Brief: 2002-1. (n.d.). Retrieved February 7, 2003, from http://www.act.org/research/briefs/2002-1.html

ACT: Research: Information Brief: 2002-2. (n.d.). Retrieved March 4, 2003, from http://www.act.org/research/briefs/2002-2.html

ACT: Standards for Transition. (n.d.). Retrieved March 4, 2003, from http://www.act.org/standard/index.html

American College Testing Program. (1996). *Mathematics for a successful transition to college: The content foundations of the ACT Assessment* (ERIC Document Reproduction Service No. ED 408 154). Iowa City, IA.

Airasian, P. W. (1997). Empiricism and values: Two faces of educational change. *International Journal of Educational Research, 27*, 433-445.

Andrade, S. (1999). *Assessing the impact of Curricular Reform-Measures of Course Efficiency and Effectiveness. AIR 1999 Annual Forum Paper.* Paper presented at the Annual Forum of the Association for Institutional Research, Seattle, WA.

Bashford, J. (2000). *How well do prerequisite courses prepare students for the next course in the sequence? Information capsule.* Report-Descriptive, Miami Dade Community College Florida.

Baxter, G. P., Shavelson, R. J., Herman, S. J., Brown, K. A., & Valadez, J. R. (1993). Mathematics performance assessment: Technical quality and diverse student impact. *Journal for Research in Mathematics Education, 24,* (3), 190-216.

Berkey, D. D., & Blanchard, P. (1995). *Calculus* (3rd ed.) Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Bontekoe, J. F. (1992). *The ACT as a predictor of college success at Trinity Christian College* (Tech. Rep. No. 143). Trinity Christian College I.. (Eric Document Reproduction Service No. ED 355 258).

Bulletin of Mississippi State University, 71. (1995, February). Mississippi State, MS: Mississippi State University, Registrar's Office.

Bulletin of Mississippi State University, 73. (1997, February). Mississippi State, MS: Mississippi State University, Registrar's Office.

Bulletin of Mississippi State University, 78. (2002, February). Mississippi State, MS: Mississippi State University, Registrar's Office.

Cizek, G. J. (1993b). Reconsidering standards and criteria. *Journal of Educational Measurement, 30* (2) 93-106.

Cizek, G. J. (1998). Filling in the blanks: Putting standardized tests to the test. *Fordham Report, 2* (11).

Cizek, G. J., Fitzgerald, S. M. & Rachor, R. E. (1995). Teacher' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3* (2), 159-179.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Duckwall, J. M., & Wilson, S. J. (1996). *Can Core Components in Mathematics Courses Replace Comprehensive Common Course Finals? AIR 1996 Annual Forum Paper.* Paper presented at the 36th Annual Forum of the Association for Institutional Research held in Albuquerque, New Mexico.

Ferrini-Mundy, J., & Graham, K. G., (1991). An overview of the calculus curriculum reform effort: Issues for learning, teaching, and curriculum development. *American Mathematical Monthly, 98* (7), 627-635.

Fisher, G. (1996). *The Validity of Pre-Calculus Multiple Choice and Performance-Based Testing as a Predictor of Undergraduate Mathematics and Chemistry Achievement.* Master's Thesis, University of California, Santa Barbara

Gainen, J. (1995). Barriers to success in quantitative gatekeeper courses. *New Directions for Teaching and Learning* (61) 5-14.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). New York: Longman Publishers USA.

Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan* 70 (9), 683.

Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership, 49* (8), 74-78.

Howell, David C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth Publishing Company.

Kohn, A. (2000). Standardized testing and its victims. *Education Week* [On-line]. Available: http://www.edweek.org/ew/ewstory.cfm?slug=04kohn.h20

Levy, L. (2001). High school grade point average and ACT mathematics sub-score as predictors of developmental mathematics students' college mathematics grade point average (Doctoral dissertation, The University of Mississippi, 2001). *UMI ProQuest Digital Dissertations,* DAI-A 62/04.

Linn, R. L., (2001). A century of standardized testing: Controversies and pendulum swings. *Educational Assessment, 7,*1

Madaus, G., & Kellaghan, T. (1993). Testing as a mechanism of public policy: A brief history and description. *Measurement & Evaluation in Counseling Development, 26* (1), 6.

Madaus, G., & O'Dwyer, L. (1999). A short history of performance assessment. 1996 *Phi Delta Kappan, 80* (9), 688.

Messick, S. (1994). *Alternative modes of assessment, uniform standards of validity.* Research report. Paper presented

Millar, S. B., et al. (1996). *A community approach to learning calculus: Fostering success for underrepresented ethnic minorities in an emerging scholars program* (Desc. Rep. No. 141). University of Wisconsin-Madison, LEAD Center.

Miller, M., & Linn, R. (2000). Validation of performance-based assessments. *Applied Psychological Measurement, 24,* (4), 367-378.

MSU General Facts and Figures (n.d.). Retrieved March 16, 2003, from http:msuinfo.ur.msstate.edu/prospect.ive/facts/general.html

MSU (2001) *Academic Operating Policy and Procedure: Academic Add/Drop Policy*: [On-line] Available: http://www.msstate.edu/dept/audit/1201.html [2003, Mar. 16].

Morgan, B. D. (1992). *The Correlation between ACT Composite Scores and Grade Point* Averages of First-Time College Freshmen after the First Year of Study at an Urban State University in Illinois (Eric Document Reproduction Service No. ED 361 342).

National Center for Fair & Open Testing (9/2/2002). *Fair Test:* [On-line] Available: http:www.fairtest.org/facts/csrtests.html

Neumann, M., (1995, April). *Graphing calculators and computers in calculus.* Report of the Calculus Committee of the Department of Mathematics and Statistics at Mississippi State University, Mississippi State, MS.

Peressini, D., & Bassett, J. (1996) Mathematical communication in students' responses to a performance-assessment task. *Communication in Mathematics, K-12 and Beyond* (pp.146-158)

Registrar's Office Tool Kit. (1995, September 15). *Announcements* (Tool Kit #6). Mississippi State University

Romagnano, L., & Long, V. (2001). The myth of objectivity in mathematics assessment. *Mathematics Teacher, 94* (1), 31.

Seese, L. (1994). *Revising the Mathematics Department.* Paper presented at the International Conference for Community College Chairs and Deans, Phoenix, AZ.

Steen, L. (1999). *Assessing Assessment.* [On-line] Available: http://www.stolaf.edu/people/steen?Papers/assessment.html.

Stevens, J. (1990). *Intermediate statistics: A modern approach.* Hillsdale. NJ: Lawrence Erlbaum Associates. Inc.

Steward, J. (1998). *Concepts and Context* (1sted.).California:Brooks/ColePublishingCompany

Steward, J. (1999). *Calculus* (3rd ed.). California: Brooks/Cole Publishing Company

Steward, J. (2000). *Calculus* (4th ed.). California: Brooks/Cole Publishing Company

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77*(3), 238.

Stiggins, R. J. (1999a). Assessment, student confidence, and school Success. *Phi Delta Kappan, 81* (3), 191.

Stiggins, R. J. (1999b). Barriers to effective student assessments. *Education Digest, 64* (6), 25.

Wilson, R. (1997). A decade of teaching "Reform *Calculus"* has been a disaster, critics charge. *Chronicle of Higher Education, 43* (22) A12-A13.

William, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative function of assessment? *British Educational Research Journal, 22* (5) p537.

APPENDIX A

MATHEMATICS STANDARDS FOR TRANSITION

EPAS Mathematics Standards for Transition by Strand and Score Range				
	Graphical Representations	**Properties of Plane Figures**	**Measurement**	**Functions †**
13–15	▪Identify the location of a point with a positive coordinate on the number line		▪Estimate or calculate the length of a line segment based on other lengths given on a geometric figure	
16–19	▪Locate points on the number line and in the first quadrant		▪Compute the perimeter of polygons when all side lengths are given ▪Compute the area of rectangles when whole number dimensions are given	
20–23	▪Comprehend the concept of length on the number line * ▪Locate points in the coordinate plane ▪Exhibit knowledge of vertical and horizontal lines and of their point of intersection ▪Exhibit knowledge of slope *	▪Exhibit knowledge of basic angle properties and special sums of angle measures (e.g., 90°, 180°, and 360°)	▪Compute the area and perimeter of triangles and rectangles in simple problems ▪Use geometric formulas when all necessary information is given	▪Work with function notation in evaluating simple quadratic functions at integer values
24–27	▪Identify the graph of a linear inequality on the number line * ▪Determine the slope of a line from points or equations * ▪Match linear graphs with their equations * ▪Find the midpoint of a line segment *	▪Use properties of isosceles triangles * ▪Recognize Pythagorean triples * ▪Use several angle properties to find an unknown angle measure	▪Compute areas and circumferences of circles after identifying necessary information ▪Compute areas of rectangles and triangles when one or more additional simple steps are required ▪Compute the perimeter of simple composite geometric figures with unknown side lengths *	▪Work with function notation in evaluating polynomial functions at integer values ▪Express the sine, cosine, and tangent of an angle in a right triangle as a ratio of given side lengths
28–32*	▪Match number line graphs with solution sets of linear inequalities ▪Use the distance formula ▪Use properties of parallel and perpendicular lines to determine an equation of a line or coordinates of a point ▪Recognize special characteristics of parabolas and circles (e.g., the vertex of a parabola and the center or radius of a circle)†	▪Apply properties of 30°-60°-90°, 45°-45°-90°, similar, and congruent triangles ▪Use the Pythagorean theorem	▪Use relationships involving area, perimeter, and volume of geometric figures to compute another measure	▪Evaluate composite functions at integer values ▪Apply basic trigonometric ratios to solve right-triangle problems
33–36†	▪Match number line graphs with solution sets of simple quadratic inequalities ▪Identify characteristics of graphs based on a set of conditions or on a general equation such as $y = ax^2 + c$ ▪Solve problems integrating multiple algebraic and/or geometric concepts	▪Draw conclusions based on a set of conditions ▪Solve multistep geometry problems that involve integrating concepts, planning, visualization, and/or making connections with other content areas ▪Use relationships among angles, arcs, and distances in a circle	▪Use scale factors to determine the magnitude of a size change ▪Compute the area of composite geometric figures when planning or visualization is required	▪Write an expression for the composite of two simple functions ▪Use trigonometric concepts and basic identities to solve problems ▪Exhibit knowledge of unit circle trigonometry ▪Match graphs of basic trigonometric functions with their equations

* PLAN & ACT Assessment only   † ACT Assessment only

Scores on EXPLORE (1–25), PLAN (1–32), and the ACT Assessment (1–36) are reported on a score scale common to all three EPAS programs.

APPENDIX B

IRB APPROVAL

Mississippi State UNIVERSITY

March 25, 2003

Maria Bearden
Mathematics & Statistics
Mail Stop 9715

Re: IRB Docket #03-052 "An Investigation of a Standardized Multiple-Choice Departmental Calculus Final Examination"

Dear Maria:

The above referenced project was reviewed and approved via administrative review on March 25, 2003 in accordance with 45 CFR 46.101 b(4). Continuing review is not necessary for this project. However, if changes are made to the protocol, you must submit a modification form for review and approval.

Please refer to your docket number (03-052) when contacting our office regarding this application

Any failure to adhere to the approved protocol could result in suspension or termination of your project. Please note that the IRB reserves the right, at anytime, to observe you and any associated researchers as they conduct the project and audit research records associated with this project.

Thank you for your cooperation and good luck to you in conducting this research project.

Sincerely yours,

Tracy S. Arwood

Tracy S. Arwood
Regulatory Compliance Officer

TSA/jm

cc: Dwight Hare
File

Office for Regulatory Compliance
P. O. Box 6223 • 300 Bowen Hall • Mailstop 9563 • Mississippi State, MS 39762 • (662) 325-3294 • FAX (662) 325-8776

APPENDIX C

SAMPLE DEPARTMENTAL FINAL EXAMINATION

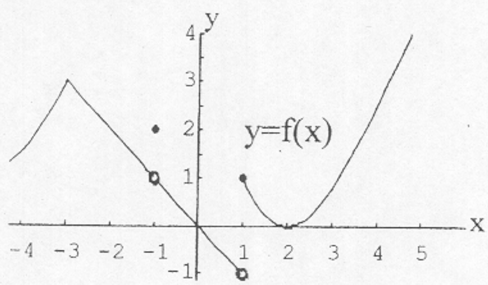
MA 1713 CALCULUS I December 7, 1996

FINAL EXAMINATION

INSTRUCTIONS:

1. Give the following information on the top of your answer sheet: YOUR NAME, I.D. NUMBER, NAME OF INSTRUCTOR, COURSE AND SECTION NUMBER, DATE.
2. Make no marks on the exam sheets. Use the scratch paper provided.
3. Use a #2 pencil to mark your answer sheet.
4. ANSWER EVERY QUESTION.
5. Be careful to mark your answer in the proper space. If you must change an answer, erase the old answer completely.
6. Return the answer sheet, the exam questions, and all scratch paper to your instructor when you finish.

Problems **1–5** refer to a function with graph



1. If $f(-1) =$
   (a) 1 (b) 2 (c) -1 (d) $f(-1)$ is not defined.

2. $\lim_{x \to 4} f(x) =$
   (a) $\frac{1}{2}$ (b) $2\frac{1}{2}$ (c) 4 (d) undefined.

3. $\lim_{x \to 1} f(x) =$
   (a) 1 (b) 2 (c) -1 (d) undefined.

4. On the interval $[-4, 5]$, the function f has discontinuities at
   (a) $x = -3$, -1 and 1 (b) $x = -1$ and 1 only
   (c) $x = 1$ only (d) $x = -3$ and 1 only.

5. On the interval $[-4, 5]$, the function f is *not* differentiable at
   (a) $x = -3$, -1 and 1 (b) $x = -1$ and 1 only
   (c) $x = 1$ only (d) $x = -3$ and 1 only.

6. If $g(x) = \begin{cases} -x & \text{if } x \le -1, \\ 4 - x^2 & \text{if } -1 < x \le 2, \\ \frac{1}{2}x - 1 & \text{if } 2 < x, \end{cases}$ then g is continuous on the intervals

(a) $(-\infty, -1]$, $(-1, 2]$, and $(2, \infty)$ only. (b) $(-\infty, -1)$, $(-1, 2)$, and $(2, \infty)$ only.

(c) $(-\infty, \infty)$. (d) $(-\infty, -1]$, and $(-1, \infty)$ only.

7. $\lim\limits_{x \to 3} \left(\dfrac{1 - x^2}{x - 2} \right)^{1/3} =$

(a) 1 (b) 2 (c) -2 (d) does not exist.

8. $\lim\limits_{x \to 0} \dfrac{\tan(3x)}{2x} =$

(a) $\frac{2}{3}$ (b) $\frac{1}{2}$ (c) $\frac{3}{2}$ (d) undefined

9. If for all $x > \sqrt{8}$, a function g satisfies $\sqrt{x^2 - 8} \le g(x) \le 4x - 11$, then $\lim\limits_{x \to 3} g(x) =$

(a) 1 (b) 0 (c) 3 (d) It's impossible to tell.

10. Evaluate $\lim\limits_{h \to 0} \dfrac{\tan\left(\frac{\pi}{4} + h\right) - \tan\left(\frac{\pi}{4}\right)}{h}$.

(a) 1 (b) 2 (c) 0 (d) The limit does not exist.

11. The slope of the line tangent to the graph of $f(x) = \dfrac{x^2}{4} + \dfrac{4}{\sqrt{x}}$ at $(4, 6)$ is

(a) 4 (b) $\frac{9}{4}$ (c) 6 (d) $\frac{7}{4}$.

12. If $y = x \cos x$, find the rate of change in y with respect to x when $x = \frac{\pi}{4}$.

(a) $\dfrac{\pi}{2\sqrt{2}}$ (b) $-\dfrac{1}{\sqrt{2}}$ (c) $\dfrac{4 - \pi}{4\sqrt{2}}$ (d) $\dfrac{4 + \pi}{\sqrt{2}}$

13. $\dfrac{d}{ds} \left(\dfrac{s + 2}{2s + 1} \right) =$

(a) $\dfrac{1}{2}$ (b) $\dfrac{-3}{(2s + 1)^2}$ (c) $\dfrac{4s + 5}{(2s + 1)^2}$ (d) $\dfrac{3 - s}{2s + 1}$

14. Find the point on the graph of $g(x) = 3 + x - \dfrac{2}{x}$, $x > 0$, where the tangent is perpendicular to the line with equation $y = 5 - \dfrac{2}{3}x$

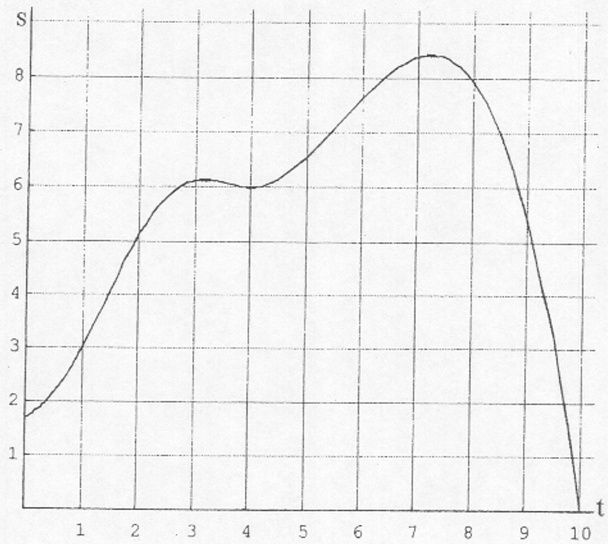
(a) $(1, 2)$ (b) $(2, 4)$ (c) $\left(\frac{1}{\sqrt{2}}, \frac{3}{\sqrt{2}}(\sqrt{2} - 1) \right)$ (d) There is no such point.

15. Suppose that f is a function with

$$f(2) = \frac{1}{2} \quad \text{and} \quad f'(2) = 4.$$

Use the line tangent to the graph of f at $(2, 1/2)$ to approximate $f(1.85)$.
(a) $f(1.85) \approx 0.1$ (b) $f(1.85) \approx -0.6$ (c) $f(1.85) \approx -0.1$ (d) $f(1.85) \approx 0.3$

Problems **16–18** refer to the motion of a particle moving along a horizontal line. Suppose that at time t (seconds) the particle is located s feet to the right of a fixed point on the line. The relation between s and t is given by the graph below.



16. The *average* velocity of the particle from time $t = 1$ to $t = 4$ seconds is
(a) 1 ft/sec (b) $\frac{3}{2}$ ft/sec (c) 3 ft/sec (d) $\frac{1}{4}$ ft/sec

17. The *instantaneous* velocity at time $t = 1$ is most nearly
(a) $\frac{1}{2}$ ft/sec (b) 2 ft/sec (c) 4 ft/sec (d) $\frac{1}{4}$ ft/sec

18. The particle is moving to the *left*
(a) always. (b) for $3 < t < 4$ and $7\frac{1}{3} < t < 10$.
(c) for $0 < t < 3$ and $4 < t < 7\frac{1}{3}$. (d) from $t = 4$ to $t = 10$.

19. If $f(x) = \sqrt{2x - x^2}$, then
(a) $f'(x) = \sqrt{2 - 2x}$ (b) $f'(x) = \dfrac{1}{2\sqrt{2 - 2x}}$
(c) $f'(x) = \dfrac{1 - x}{\sqrt{2x - x^2}}$ (d) $f'(x) = \dfrac{1}{2\sqrt{2x - x^2}}$.

20. Suppose that f and g are differentiable functions with

$$\begin{array}{ll} g(0) = -3 & g'(0) = 7 \\ f(3) = 5 & f(7) = -1 \\ f'(-3) = 2 & f'(7) = 10. \end{array}$$

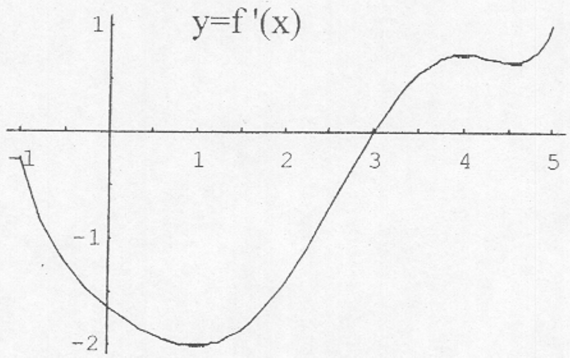
If $h(x) = f(g(x))$, then $h'(0) =$
(a) 14 (b) 2 (c) 10 (d) 35

21. If $f(x) = (3x-10)(x+2)^3$, then the graph of f has horizontal tangent at $(x, f(x))$ for $x =$
(a) $\frac{10}{3}$ and -2 only (b) -2 only (c) -2 and 2 only (d) f has no horizontal tangents

22. If $f(x) = \dfrac{x+3}{x+9}$, find $f''(-5)$
(a) 1 (b) $\frac{1}{32}$ (c) 0 (d) $\frac{-3}{16}$

23. If $f(x) = \sin x$, then $f^{(161)}(x) =$
(a) $\sin x$ (b) $\cos x$ (c) $-\sin x$ (d) $-\cos x$.

Problems **24** and **25** refer to a function f with *derivative* graphed below



24. The function f is *increasing* on the interval(s)
(a) $(-1, 3)$ (b) $(-1, 1)$ and $(4, 4\frac{1}{2})$ (c) $(3, 5)$ (d) $(1, 4)$ and $(4\frac{1}{2}, 5)$

25. The function f has a relative extremum at $x =$
(a) 1 and $4\frac{1}{2}$ (b) 4 (c) 3 (d) 1 and 5

Problems **26** and **27** refer to a function g with *derivative*

$$g'(x) = 2x^3 - x^4.$$

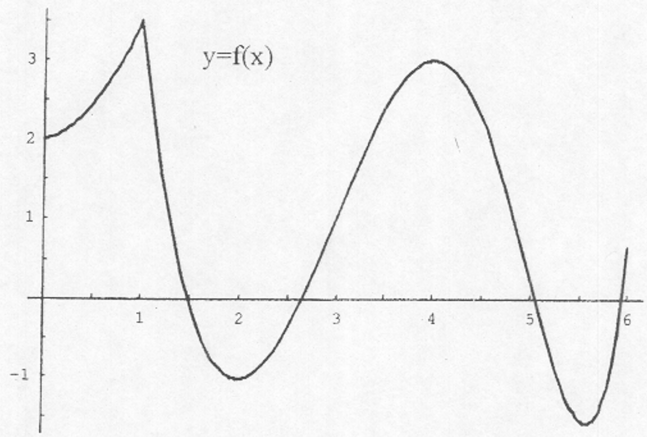
26. The function g has a relative maximum at $x =$
(a) 0 only (b) 2 only (c) 0 and 2 (d) g has no relative maximum.

27. g is concave *upward* on the interval
(a) $\left(\frac{3}{2},\infty\right)$ only (b) $(-\infty,0)$ and $(2,\infty)$ only

(c) $\left(-\infty,\frac{3}{2}\right)$ only (d) $(-\infty,0)$ and $\left(\frac{3}{2},\infty\right)$ only

28. Suppose that y is a differentiable function of x satisfying $\dfrac{dy}{dx}=\dfrac{1}{x+y}$. Find $\dfrac{d^2y}{dx^2}$
(a) $\dfrac{x+y}{x+y+1}$ (b) $-\dfrac{1}{(x+y)^3}$ (c) $-\dfrac{x+y+1}{(x+y)^3}$ (d) $-\dfrac{1}{(x+y)^2}$

Problems **29** and **30** refer to the graph of a function $y=f(x)$ below.



29. The absolute maximum and minimum values of f on the interval $[0,5]$ are respectively
(a) $3\frac{1}{2}$ and $-1\frac{1}{2}$ (b) 3 and -1 (c) 2 and 5 (d) $3\frac{1}{2}$ and -1

30. The graph of f is concave *downwards* on the interval(s)
(a) $(1,2)$ and $\left(4,5\frac{1}{2}\right)$ (b) $(0,3)$ and $(5,6)$ (c) $(3,5)$ (d) $(1,4)$

31. If g is a function such that $g'(7)=0$ and $g''(x)=4(x+1)(x-10)$, then
(a) $(7,g(7))$ is a relative maximum (b) $(7,g(7))$ is a relative minimum

(c) $(7,g(7))$ is an inflection point (d) no conclusion is possible.

32. Find all numbers c satisfying the conclusion of the Mean Value Theorem for the function $f(x)=\frac{1}{2}x-\sqrt{1-x^2}$ on the interval $[-1,1]$.
(a) $c=0$ only (b) $c=\frac{1}{2}$ only (c) $c=\pm\frac{1}{\sqrt{2}}$
(d) Since f is not differentiable at ± 1, the Mean Value Theorem does not apply.

33. The graph of $y = \dfrac{3x - 3x^3}{x^3 + x^2}$ has asymptotes

(a) $x = 0$, $x = -1$ and $y = -3$ only (b) $x = 0$ and $y = 0$ only

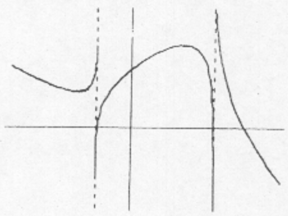
(c) $x = -1$ and $y = 0$ only (d) $x = 0$ and $y = -3$ only

34. $\displaystyle\lim_{x \to -\infty} \frac{4 - 3x}{\sqrt{1 + 9x^2}} =$

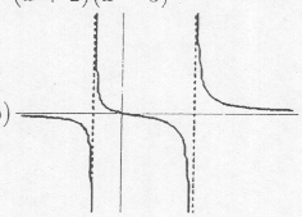
(a) 1 (b) $\frac{4}{3}$ (c) 0 (d) undefined

35. The graph of the function $f(x) = \dfrac{x}{(x+2)(x-5)}$ most closely resembles

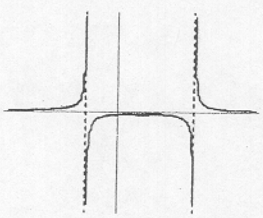
(a)



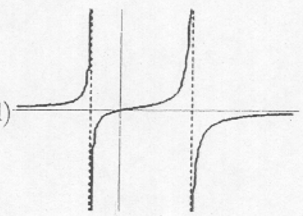
(b)



(c)



(d)

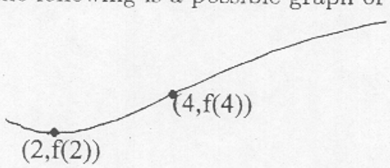


36. Suppose that a function f satisfies

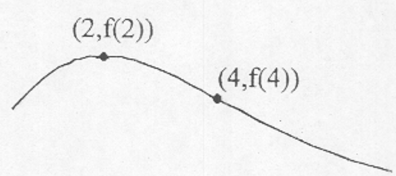
- f is differentiable on $(0, 6)$.
- $f'(x) < 0$ if $x < 2$, and $f'(x) > 0$ if $x > 2$.
- f has an inflection point at $(4, f(4))$.

Which of the following is a possible graph of f ?

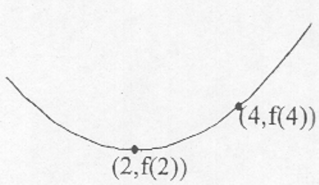
(a)



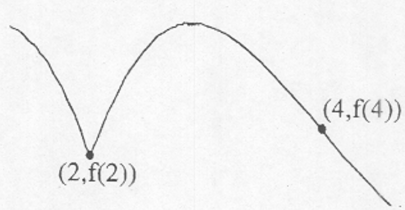
(b)



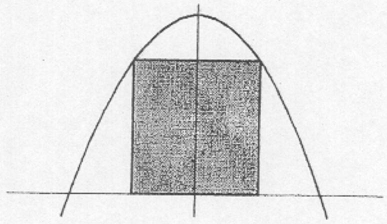
(c)



(d)



37. Find the greatest area of a rectangle with two vertices on the x-axis and two on the graph of $y = 4 - x^2$.

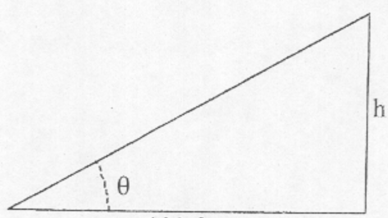


(a) $\frac{21}{4} = 5.25$ (b) $\frac{32\sqrt{3}}{9} \approx 6.16$ (c) $\frac{8\sqrt{5}}{3} \approx 5.96$ (d) $\frac{8\sqrt{6}}{3} \approx 6.53$

38. An apple orchard has 25 trees per acre, and the average yield is 300 apples per tree. For each additional tree planted per acre, it's estimated that the average yield will decline by 7 apples per tree. How many trees per acre will yield the greatest crop? (Of course, no grower would plant a fraction of a tree.)

(a) 28 (b) 22 (c) 38 (d) 34

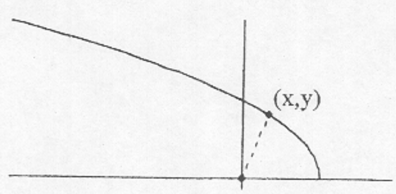
39. A right triangle has height $h = h(t)$ and base of constant length 100 feet. If h is increasing at a rate of 15 ft/sec, find the rate of change in θ at the instant $h = 100$ feet.



(a) .025 radians/sec $\approx 1.4°$/sec (b) .125 radians/sec $\approx 7.1°$/sec

(c) .075 radians/sec $\approx 4.3°$/sec (d) .20 radians/sec $\approx 11.5°$/sec

40. Let c be a constant, $c > \frac{1}{2}$. Find the point on the graph of $y = \sqrt{c - x}$ closest to the origin.



(a) $\left(\frac{c}{2}, \sqrt{\frac{c}{2}}\right)$ (b) $(c, 0)$ (c) $\left(\frac{c}{\sqrt{2}}, \frac{\sqrt{c}}{\sqrt{2(\sqrt{2}+1)}}\right)$ (d) none of these

APPENDIX D

FREQUENCY AND PERCENT TABLES OF DATA FROM I.T.S.

Table 1

Frequency and Percent of Calculus I Grades with SMCDF (DEPT group)

DEPT	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	135	12.7	17.7	17.7
	B	219	20.6	28.7	46.4
	C	216	20.3	28.3	74.7
	D	101	9.5	13.3	88.0
	F	91	8.6	11.9	100.0
	Total	762	71.7	100.0	
Missing	S	80	7.5		
	AU	193	18.2		
	W	28	2.6		
	Total	301	28.3		
Total		1063	100.0		

Table 2

Frequency and Percent of Calculus II Grades with SMCDF (DEPT group)

DEPT	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	119	11.2	24.6	24.6
	B	125	11.8	25.9	50.5
	C	118	11.1	24.4	74.9
	D	48	4.5	9.9	84.8
	F	73	6.9	15.1	100.0
	Total	483	45.4	100.0	
Missing	DNE	443	41.7		
Missing	S	3	.3		
	AU	106	10.0		
	W	28	2.6		
	Total	580	54.6		
Total		1063	100.0		

Table 3

Frequency and Percent of Calculus III Grades with SMCDF (DEPT group)

DEPT	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	80	7.5	20.6	20.6
	B	96	9.0	24.7	45.3
	C	109	10.3	28.1	73.4
	D	54	5.1	13.9	87.3
	F	49	4.6	12.6	100.0
	Total	388	36.5	100.0	
Missing	DNE	617	58.0		
	AU	42	4.0		
	W	16	1.5		
	Total	675	63.5		
Total		1063	100.0		

Table 4

Frequency and Percent of Calculus IV Grades with SMCDF (DEPT group)

DEPT	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	108	10.2	31.7	31.7
	B	103	9.7	30.2	61.9
	C	72	6.8	21.1	83.0
	D	33	3.1	9.7	92.7
	F	25	2.4	7.3	100.0
	Total	341	32.1	100.0	
Missing	DNE	696	65.5		
	AU	5	.5		
	W	20	1.9		
	System	1	.1		
	Total	722	67.9		
Total		1063	100.0		

Table 5

Frequency and Percent of Calculus I Grades without SMCDF (INST group)

INST	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	221	14.8	18.9	18.9
	B	222	14.8	19.0	37.9
	C	264	17.6	22.6	60.5
	D	162	10.8	13.9	74.4
	F	299	20.0	25.6	100.0
	Total	1168	78.0	100.0	
Missing	S	182	12.2		
	AU	2	.1		
	W	145	9.7		
	Total	329	22.0		
Total		1497	100.0		

Table 6

Frequency and Percent of Calculus II Grades without SMCDF (INST group)

INST	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	149	10.0	23.2	23.2
	B	130	8.7	20.2	43.4
	C	163	10.9	25.3	68.7
	D	72	4.8	11.2	79.9
	F	129	8.6	20.1	100.0
	Total	643	43.0	100.0	
Missing	DNE	783	52.3		
Missing	S	5	.3		
	AU	2	.1		
	W	64	4.3		
	Total	854	57.0		
Total		1497	100.0		

Table 7

Frequency and Percent of Calculus III Grades without SMCDF (INST group)

INST	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	116	7.7	25.7	25.7
	B	130	8.7	28.8	54.5
	C	80	5.3	17.7	72.2
	D	54	3.6	11.9	84.1
	F	72	4.8	15.9	100.0
	Total	452	30.2	100.0	
Missing	DNE	1005	67.1		
	W	40	2.7		
	Total	1045	69.8		
Total		1497	100.0		

Table 8

Frequency and Percent of Calculus IV grades without SMCDF (INST group)

INST	Grade	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	130	8.7	33.0	33.0
	B	125	8.4	31.7	64.7
	C	71	4.7	18.0	82.7
	D	46	3.1	11.7	94.4
	F	22	1.5	5.6	100.0
	Total	394	26.3	100.0	
Missing	DNE	1080	72.1		
	W	23	1.5		
	Total	1103	73.7		
Total		1497	100.0		

Table 9

Frequency and Percent of ACT Mathematics scores

DEPT	ACT	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15	6	.6	.6	.6
	16	7	.7	.7	1.4
	17	16	1.5	1.7	3.1
	18	41	3.9	4.3	7.4
	19	37	3.5	3.9	11.3
	20	43	4.0	4.6	15.9
	21	49	4.6	5.2	21.1
	22	52	4.9	5.5	26.6
	23	71	6.7	7.5	34.1
	24	47	4.4	5.0	39.1
	25	97	9.1	10.3	49.4
	26	70	6.6	7.4	56.8
	27	86	8.1	9.1	65.9
	28	80	7.5	8.5	74.4
	29	71	6.7	7.5	81.9
	30	56	5.3	5.9	87.8
	31	51	4.8	5.4	93.2
	32	28	2.6	3.0	96.2
	33	19	1.8	2.0	98.2
	34	12	1.1	1.3	99.5
	36	5	.5	.5	100.0
	Total	944	88.8	100.0	
Missing	System	119	11.2		
Total		1063	100.0		

Table 10

Frequency and Percent of ACT Mathematics scores

INST	ACT	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	14	2	.1	.1	.1
	15	9	.6	.7	.8
	16	20	1.3	1.4	2.2
	17	39	2.6	2.8	5.1
	18	49	3.3	3.5	8.6
	19	51	3.4	3.7	12.3
	20	68	4.5	4.9	17.2
	21	73	4.9	5.3	22.5
	22	55	3.7	4.0	26.5
	23	76	5.1	5.5	32.0
	24	78	5.2	5.6	37.6
	25	98	6.5	7.1	44.7
	26	106	7.1	7.7	52.3
	27	128	8.6	9.3	61.6
	28	108	7.2	7.8	69.4
	29	94	6.3	6.8	76.2
	30	95	6.3	6.9	83.1
	31	87	5.8	6.3	89.4
	32	85	5.7	6.1	95.5
	33	21	1.4	1.5	97.0
	34	20	1.3	1.4	98.5
	35	6	.4	.4	98.9
	36	15	1.0	1.1	100.0
	Total	1383	92.4	100.0	
Missing	System	114	7.6		
Total		1497	100.0		

Table 11

Frequency and Percent of DEPT Group ACT Mathematics Scores by Ranges

DEPT	Frequency	Percent	Valid Percent	Cumulative Percent
ACTm<=19	107	10.1	11.3	11.3
ACTm=20-23	215	20.2	22.8	34.1
ACTm=24-27	300	28.2	31.8	65.9
ACTm=28-32	286	26.9	30.3	96.2
ACTm=33-36	36	3.4	3.8	100.0
Total	944	88.8	100.0	
Missing System	119	11.2		
Total	1063	100.0		

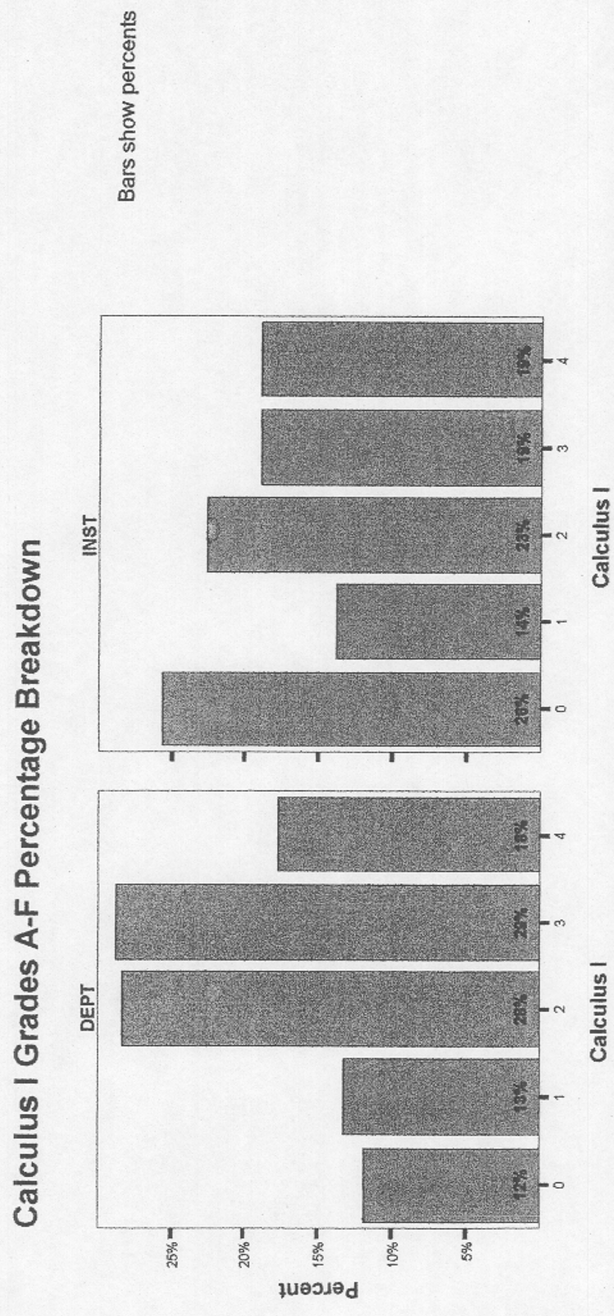
Table 12

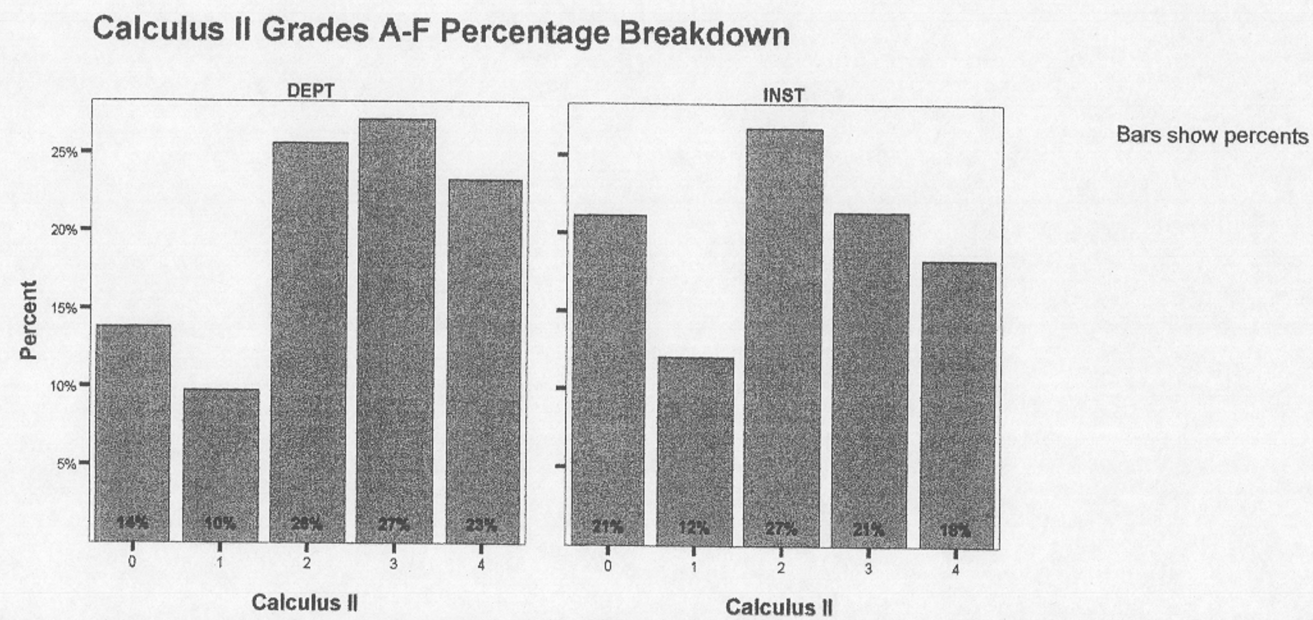
Frequency and Percent of INST Group ACT Mathematics Scores by Ranges

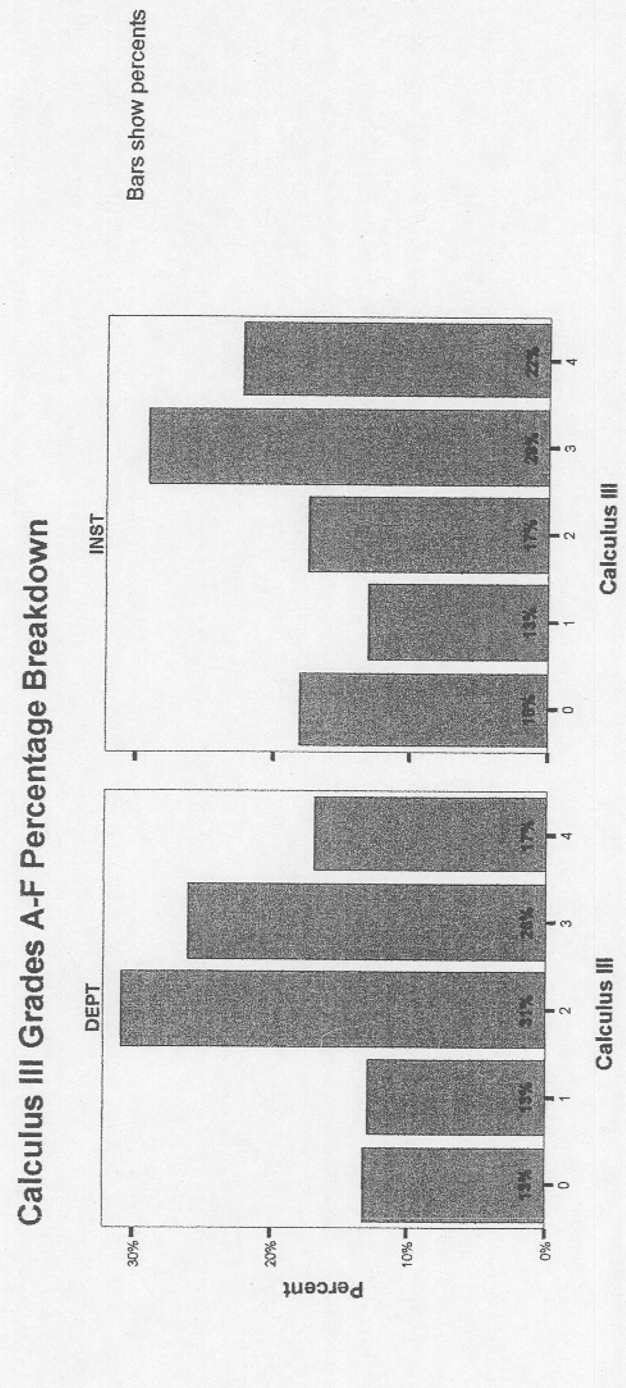
INST	Frequency	Percent	Valid Percent	Cumulative Percent
ACTm<=19	170	11.4	12.3	12.3
ACTm=20-23	272	18.2	19.7	32.0
ACTm=24-27	410	27.4	29.6	61.6
ACTm=28-32	469	31.3	33.9	95.5
ACTm=33-36	62	4.1	4.5	100.0
Total	1383	92.4	100.0	
Missing System	114	7.6		
Total	1497	100.0		

APPENDIX E

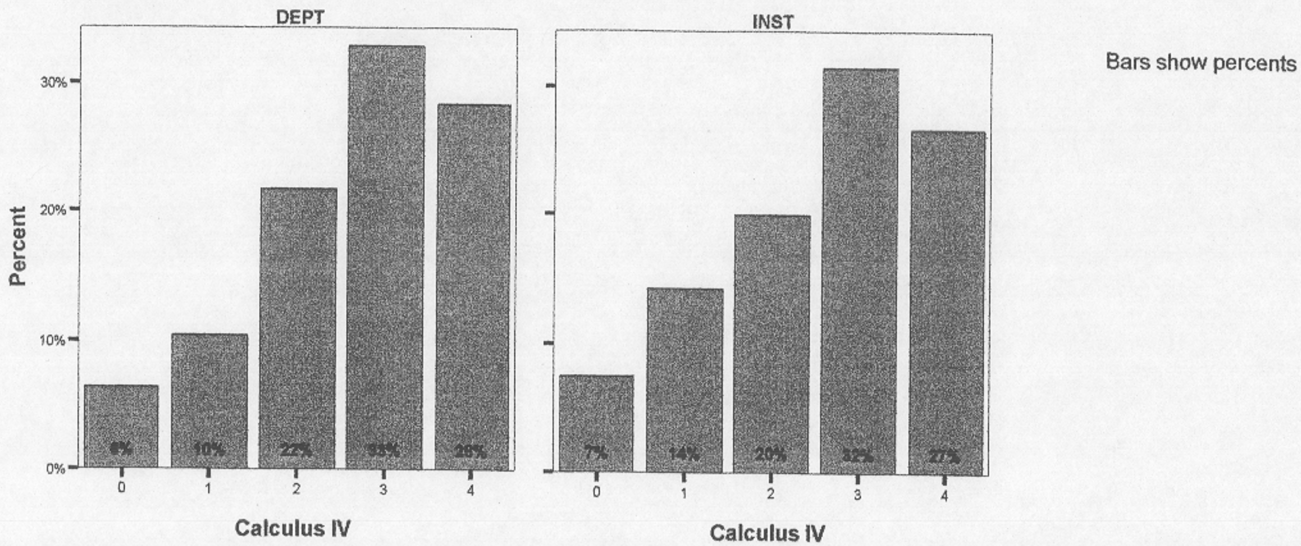
CALCULUS I-IV GRADE DISTRIBUTION BAR GRAPHS


Calculus I Grades A-F Percentage Breakdown
DEPT
INST
Percent
25%
20%
15%
10%
5%
12%
13%
28%
29%
18%
26%
14%
23%
19%
19%
0
1
2
3
4
Calculus I
Calculus I
Bars show percents


Calculus II Grades A-F Percentage Breakdown
DEPT
INST
Percent
25%
20%
15%
10%
5%
14%
10%
26%
27%
23%
21%
12%
27%
21%
18%
0
1
2
3
4
Calculus II
Calculus II
Bars show percents


Calculus III Grades A-F Percentage Breakdown
DEPT
INST
Percent
30%
20%
10%
0%
0
1
2
3
4
Calculus III
Calculus III
Bars show percents

Calculus IV Grades A-F Percentage Breakdown

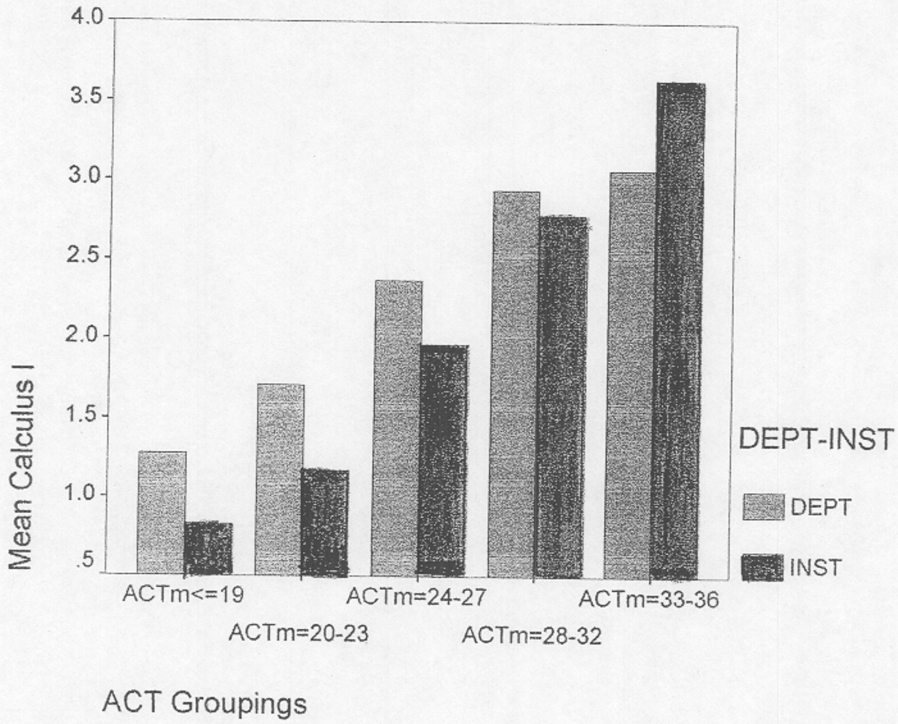


APPENDIX F

CALCULUS I-IV GROUP MEANS BY ACT RANGES

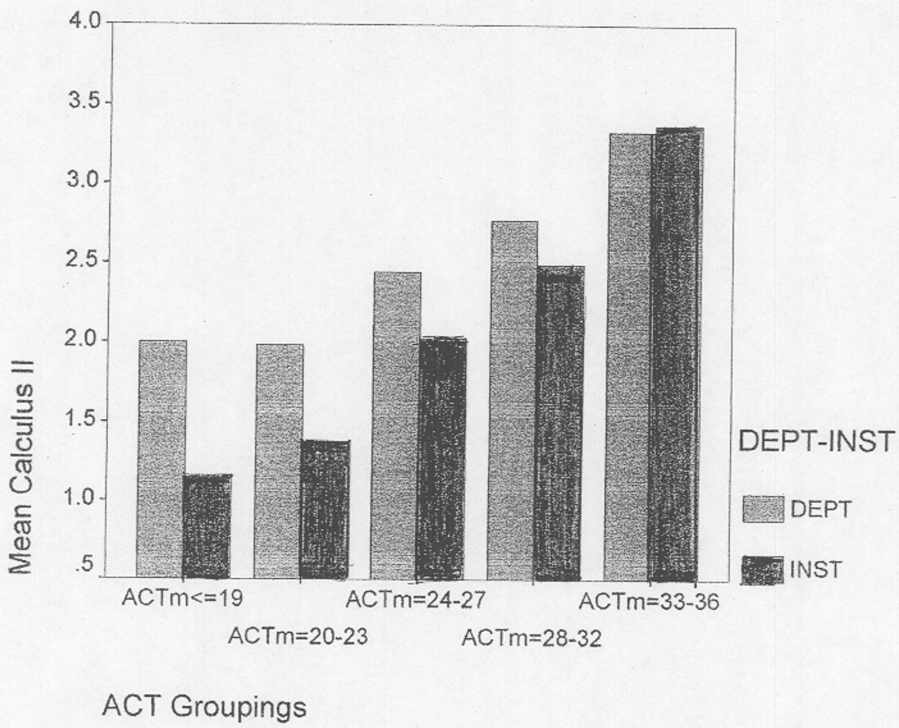
Calculus I Means by ACT Mathematics for DEPT and INST Groups

Graph



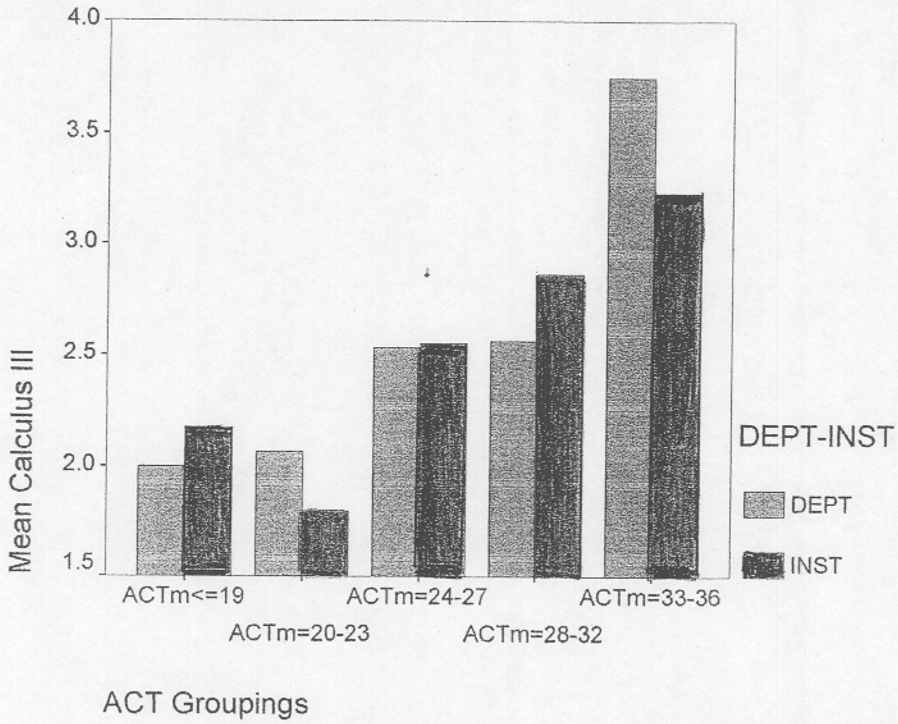
Calculus II Means by ACT Mathematics for DEPT and INST Groups

Graph



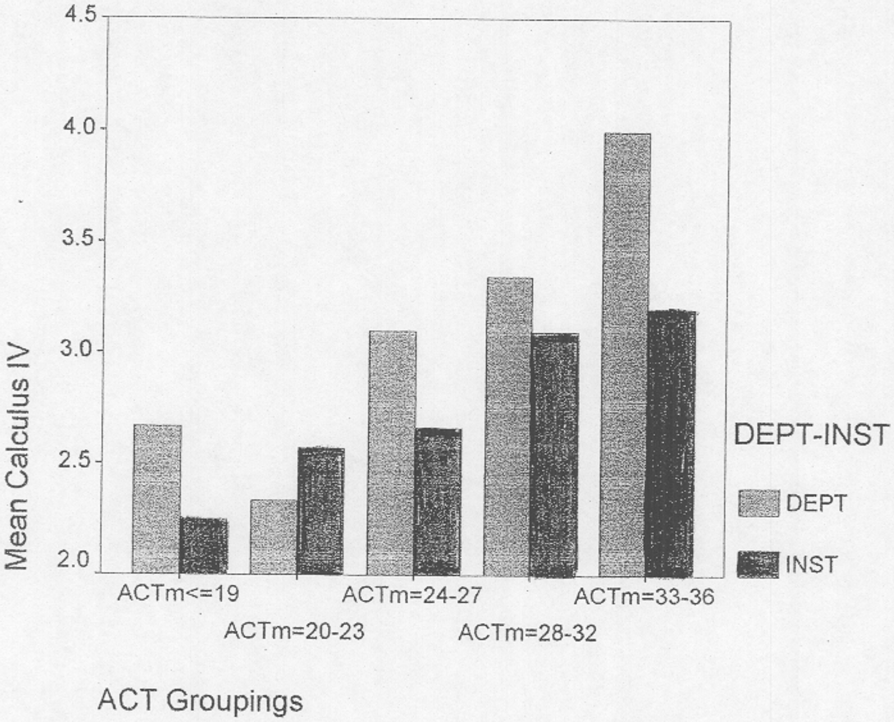
Calculus III Means by ACT Mathematics for DEPT and INST Groups

Graph



Calculus IV Means by ACT Mathematics for DEPT and INST Groups

Graph



APPENDIX G

CALCULUS II-IV ANOVA STATISTICS AND INTERACTION PLOTS

Calculus II ANOVA Statistics and Interaction Plots

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
ACT Groupings	1	ACTm<=19	38
	2	ACTm=20-23	113
	3	ACTm=24-27	261
	4	ACTm=28-32	301
	5	ACTm=33-36	25
DEPT-INST	0	DEPT	310
	1	INST	428

Calculus II ANOVA Statistics and Interaction Plots

Tests of Between-Subjects Effects

Dependent Variable: Calculus II

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	166.358[b]	9	18.484	11.573	.000	104.158	1.000
Intercept	1438.914	1	1438.914	900.913	.000	900.913	1.000
ACT_GRP	118.277	4	29.569	18.513	.000	74.054	1.000
DEPTINST	12.174	1	12.174	7.622	.006	7.622	.787
ACT_GRP * DEPTINST	4.962	4	1.241	.777	.540	3.107	.251
Error	1162.742	728	1.597				
Total	5190.000	738					
Corrected Total	1329.100	737					

a. Computed using alpha = .05

b. R Squared = .125 (Adjusted R Squared = .114)

Profile Plots

Calculus II ANOVA Statistics and Interaction Plots

Descriptive Statistics

Dependent Variable: Calculus II

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.00	1.183	11
	INST	1.15	1.292	27
	Total	1.39	1.306	38
ACTm=20-23	DEPT	1.98	1.254	55
	INST	1.38	1.400	58
	Total	1.67	1.359	113
ACTm=24-27	DEPT	2.45	1.328	116
	INST	2.03	1.301	145
	Total	2.21	1.327	261
ACTm=28-32	DEPT	2.77	1.093	119
	INST	2.49	1.278	182
	Total	2.60	1.214	301
ACTm=33-36	DEPT	3.33	1.323	9
	INST	3.38	.885	16
	Total	3.36	1.036	25
Total	DEPT	2.50	1.259	310
	INST	2.13	1.382	428
	Total	2.29	1.343	738

Levene's Test of Equality of Error Variances[a]

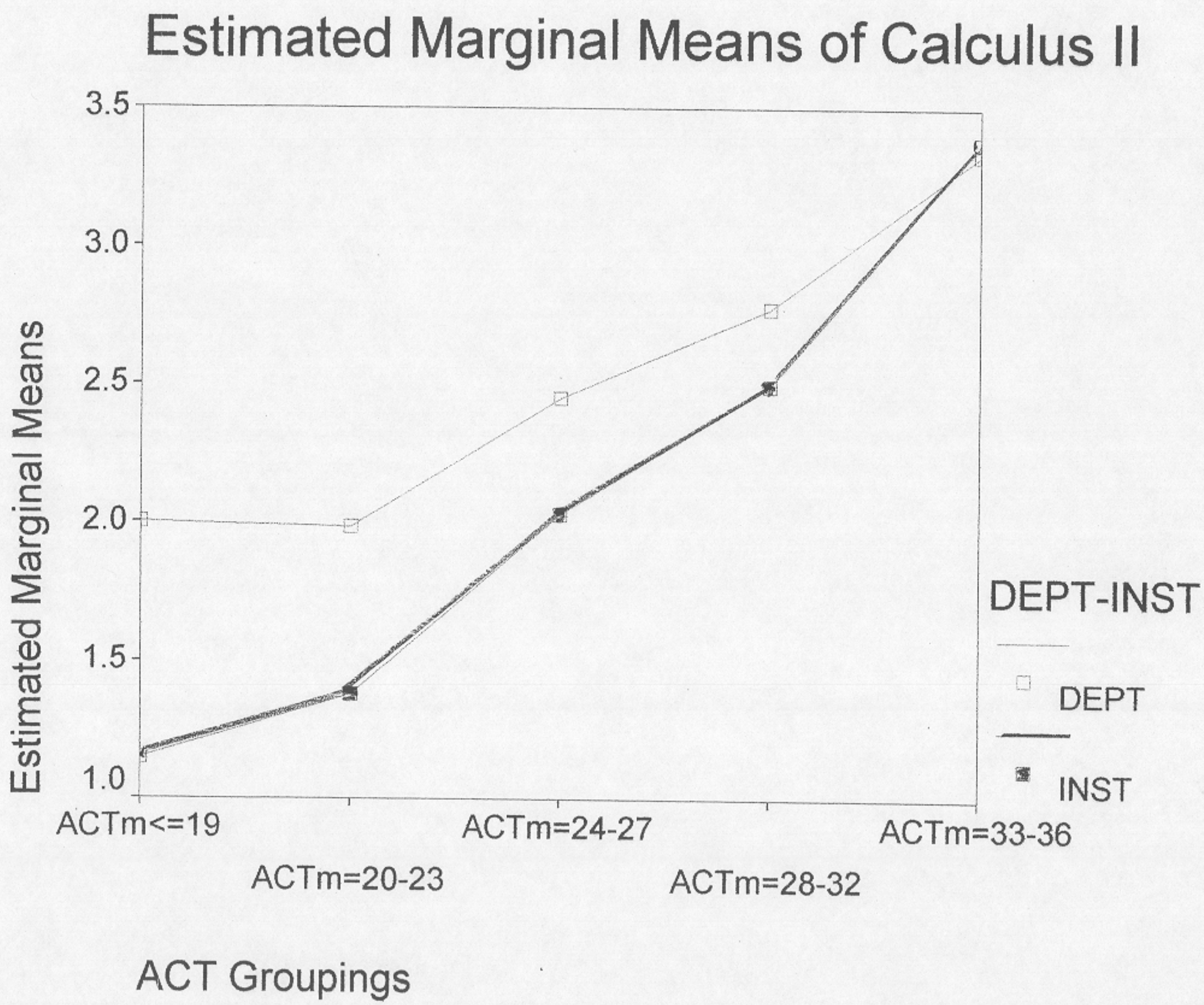
Dependent Variable: Calculus II

F	df1	df2	Sig.
2.010	9	728	.036

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+ACT_GRP+DEPTINST+ACT_GRP * DEPTINST

Calculus II ANOVA Statistics and Interaction Plots



Calculus III ANOVA Statistics and Interaction Plots

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
ACT Groupings	1	ACTm<=19	14
	2	ACTm=20-23	50
	3	ACTm=24-27	138
	4	ACTm=28-32	167
	5	ACTm=33-36	17
DEPT-INST	0	DEPT	175
	1	INST	211

Calculus III ANOVA Statistics and Interaction Plots

Descriptive Statistics

Dependent Variable: Calculus III

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.00	1.414	8
	INST	2.17	.983	6
	Total	2.07	1.207	14
ACTm=20-23	DEPT	2.07	1.172	30
	INST	1.80	1.508	20
	Total	1.96	1.309	50
ACTm=24-27	DEPT	2.53	1.112	62
	INST	2.55	1.248	76
	Total	2.54	1.185	138
ACTm=28-32	DEPT	2.56	1.156	71
	INST	2.86	1.184	96
	Total	2.74	1.178	167
ACTm=33-36	DEPT	3.75	.500	4
	INST	3.23	1.301	13
	Total	3.35	1.169	17
Total	DEPT	2.47	1.168	175
	INST	2.65	1.279	211
	Total	2.57	1.232	386

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Calculus III

F	df1	df2	Sig.
1.132	9	376	.339

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+ACT_GRP+DEPTINST+ACT_GRP * DEPTINST

Calculus III ANOVA Statistics and Interaction Plots

Tests of Between-Subjects Effects

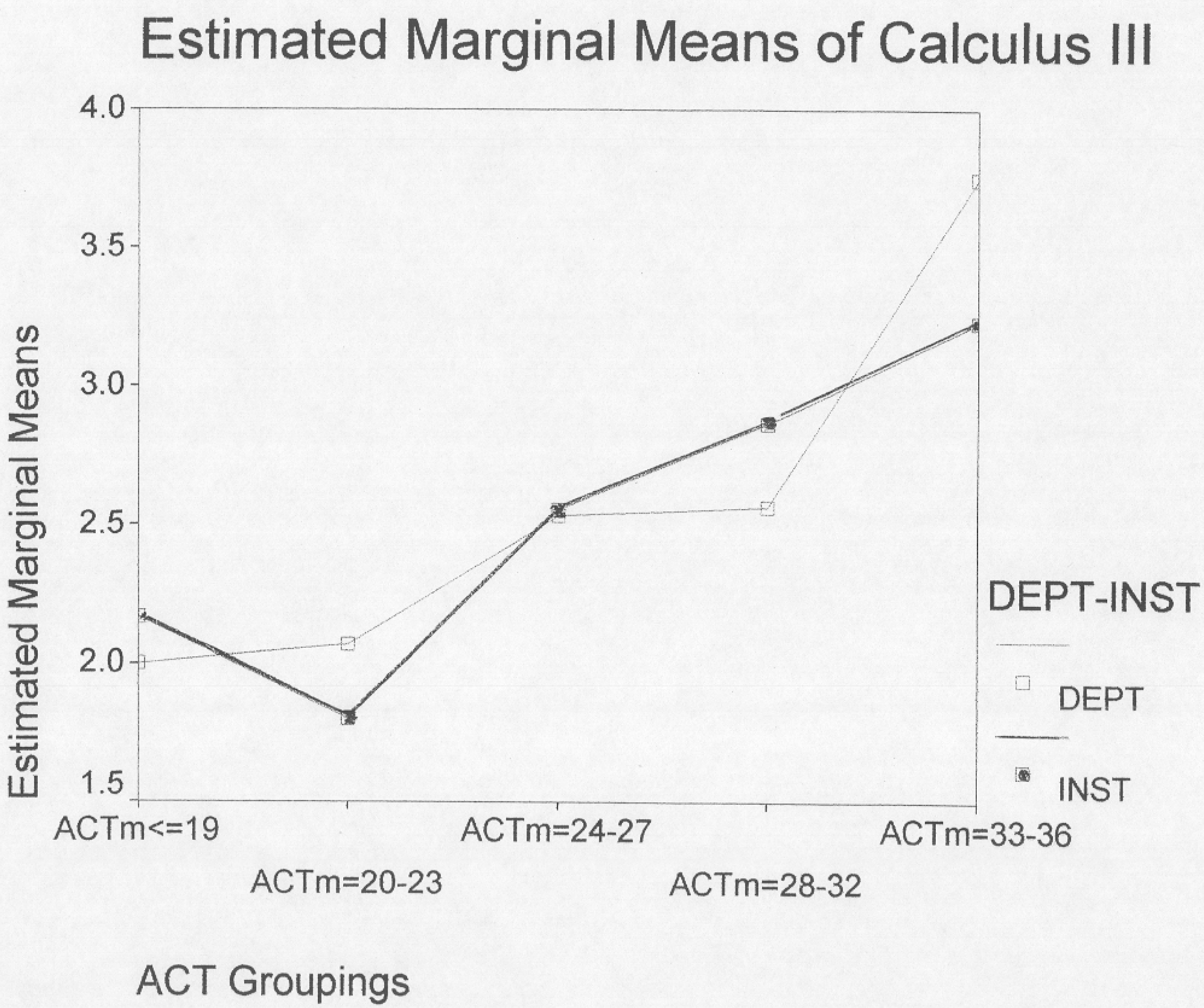
Dependent Variable: Calculus III

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	42.724[b]	9	4.747	3.294	.001	29.645	.983
Intercept	862.268	1	862.268	598.303	.000	598.303	1.000
ACT_GRP	36.454	4	9.113	6.324	.000	25.294	.989
DEPTINST	.117	1	.117	.081	.776	.081	.059
ACT_GRP * DEPTINST	4.667	4	1.167	.810	.520	3.238	.259
Error	541.887	376	1.441				
Total	3134.000	386					
Corrected Total	584.611	385					

a. Computed using alpha = .05

b. R Squared = .073 (Adjusted R Squared = .051)

Profile Plots


Estimated Marginal Means of Calculus III
Estimated Marginal Means
4.0
3.5
3.0
2.5
2.0
1.5
ACTm<=19
ACTm=20-23
ACTm=24-27
ACTm=28-32
ACTm=33-36
ACT Groupings
DEPT-INST
DEPT
INST

Calculus IV ANOVA Statistics and Interaction Plots

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
ACT Groupings	1	ACTm<=19	7
	2	ACTm=20-23	25
	3	ACTm=24-27	97
	4	ACTm=28-32	128
	5	ACTm=33-36	14
DEPT-INST	0	DEPT	121
	1	INST	150

Calculus IV ANOVA Statistics and Interaction Plots

Descriptive Statistics

Dependent Variable: Calculus IV

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.67	.577	3
	INST	2.25	1.500	4
	Total	2.43	1.134	7
ACTm=20-23	DEPT	2.33	1.455	18
	INST	2.57	1.134	7
	Total	2.40	1.354	25
ACTm=24-27	DEPT	3.10	.889	41
	INST	2.66	1.066	56
	Total	2.85	1.014	97
ACTm=28-32	DEPT	3.35	.821	55
	INST	3.10	1.095	73
	Total	3.20	.991	128
ACTm=33-36	DEPT	4.00	.000	4
	INST	3.20	.789	10
	Total	3.43	.756	14
Total	DEPT	3.12	1.010	121
	INST	2.89	1.094	150
	Total	2.99	1.061	271

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Calculus IV

F	df1	df2	Sig.
3.470	9	261	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+ACT_GRP+DEPTINST+ACT_GRP * DEPTINST

Calculus IV ANOVA Statistics and Interaction Plots

Tests of Between-Subjects Effects

Dependent Variable: Calculus IV

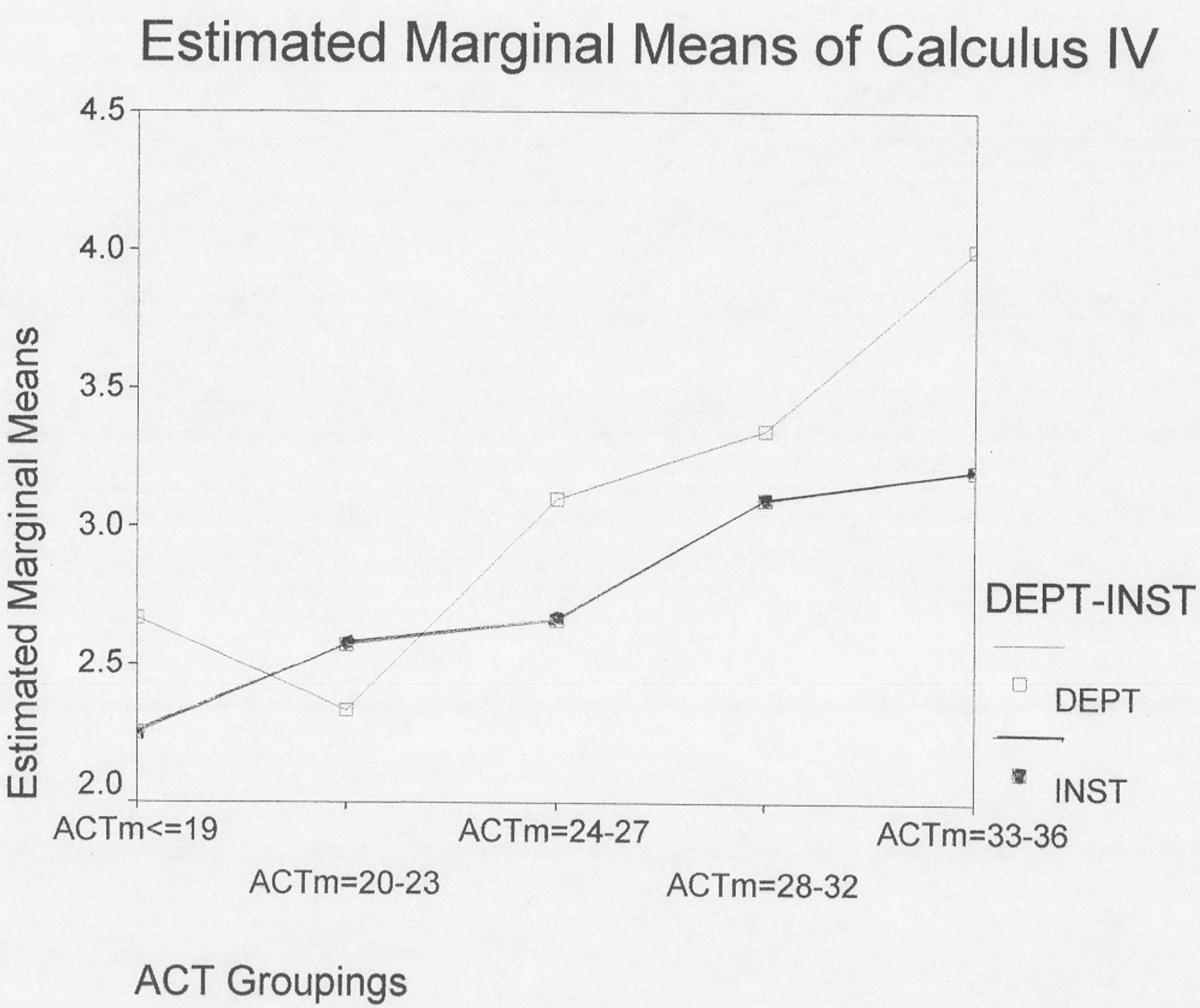
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	30.326[b]	9	3.370	3.214	.001	28.923	.979
Intercept	708.092	1	708.092	675.336	.000	675.336	1.000
ACT_GRP	19.395	4	4.849	4.625	.001	18.498	.945
DEPTINST	2.299	1	2.299	2.193	.140	2.193	.314
ACT_GRP * DEPTINST	2.715	4	.679	.647	.629	2.589	.210
Error	273.659	261	1.049				
Total	2731.000	271					
Corrected Total	303.985	270					

a. Computed using alpha = .05

b. R Squared = .100 (Adjusted R Squared = .069)

Profile Plots

Calculus IV ANOVA Statistics and Interaction Plots



APPENDIX H

CALCULUS II-IV ANOVA STATISTICS AND INTERACTION PLOTS
(Excluding F Grades)

Statistics for Calculus II Using Grades A through D.

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
ACT Groupings	1	ACTm<=19	24
	2	ACTm=20-23	77
	3	ACTm=24-27	219
	4	ACTm=28-32	280
	5	ACTm=33-36	24
DEPT-INST	0	DEPT	277
	1	INST	347

Statistics for Calculus II Using Grades A through D.

Descriptive Statistics

Dependent Variable: Calculus II

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.44	.726	9
	INST	2.07	1.033	15
	Total	2.21	.932	24
ACTm=20-23	DEPT	2.53	.767	43
	INST	2.35	1.012	34
	Total	2.45	.882	77
ACTm=24-27	DEPT	2.78	1.030	102
	INST	2.51	.934	117
	Total	2.64	.987	219
ACTm=28-32	DEPT	2.87	.978	115
	INST	2.75	1.044	165
	Total	2.80	1.017	280
ACTm=33-36	DEPT	3.75	.463	8
	INST	3.38	.885	16
	Total	3.50	.780	24
Total	DEPT	2.80	.968	277
	INST	2.63	1.021	347
	Total	2.71	1.001	624

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Calculus II

F	df1	df2	Sig.
2.371	9	614	.012

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+ACT_GRP+DEPTINST+ACT_GRP * DEPTINST

Statistics for Calculus II Using Grades A through D.

Tests of Between-Subjects Effects

Dependent Variable: Calculus II

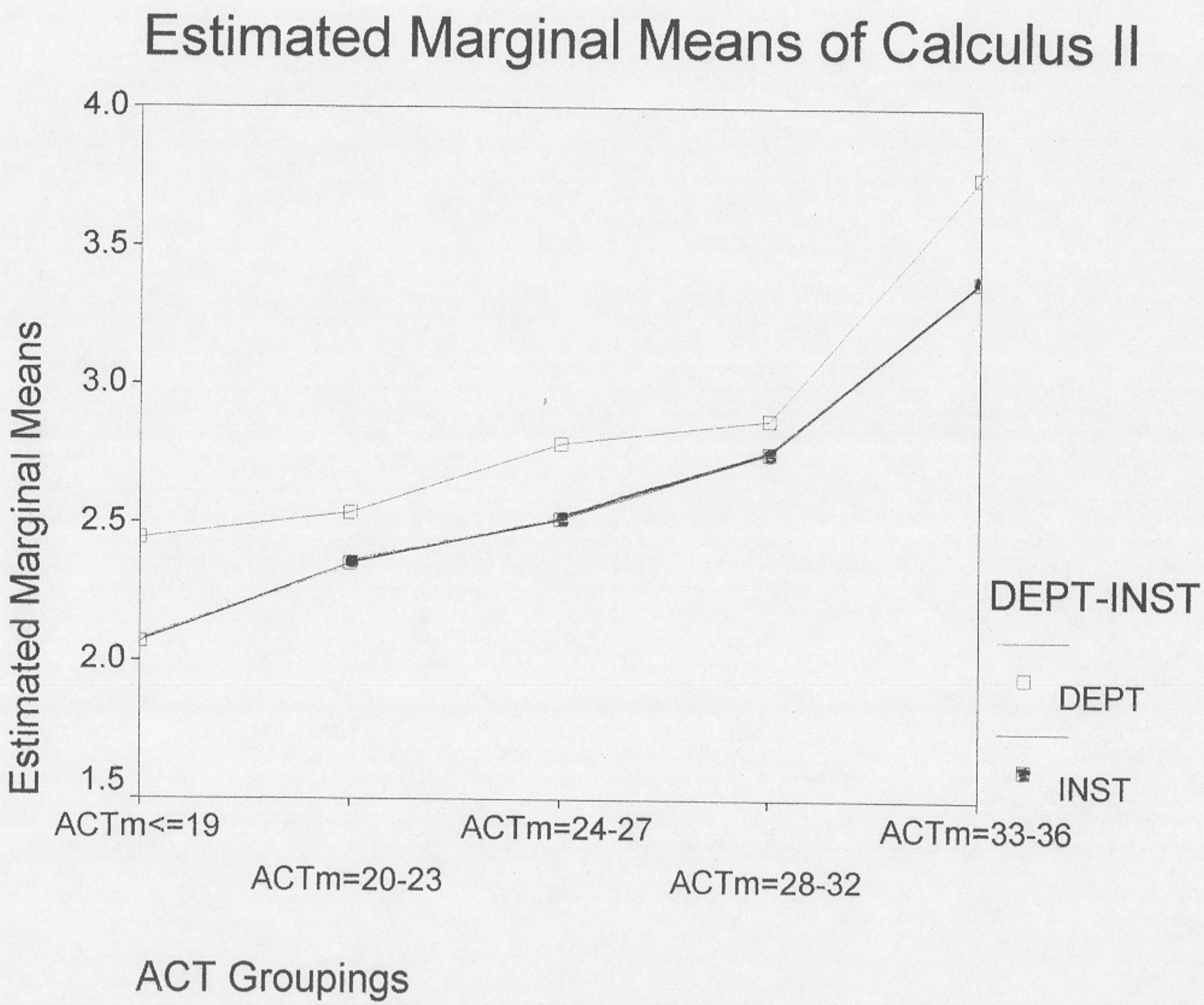
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	36.534[b]	9	4.059	4.245	.000	38.201	.998
Intercept	1669.587	1	1669.587	1745.761	.000	1745.761	1.000
ACT_GRP	29.087	4	7.272	7.603	.000	30.414	.997
DEPTINST	3.888	1	3.888	4.065	.044	4.065	.521
ACT_GRP * DEPTINST	1.081	4	.270	.283	.889	1.130	.113
Error	587.209	614	.956				
Total	5190.000	624					
Corrected Total	623.744	623					

a. Computed using alpha = .05

b. R Squared = .059 (Adjusted R Squared = .045)

Profile Plots

Statistics for Calculus II Using Grades A through D.



Statistics for Calculus III Using Grades A through D.

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
ACT Groupings	1	ACTm<=19	13
	2	ACTm=20-23	41
	3	ACTm=24-27	127
	4	ACTm=28-32	153
	5	ACTm=33-36	16
DEPT-INST	0	DEPT	161
	1	INST	189

Statistics for Calculus III Using Grades A through D.

Descriptive Statistics

Dependent Variable: Calculus III

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.29	1.254	7
	INST	2.17	.983	6
	Total	2.23	1.092	13
ACTm=20-23	DEPT	2.30	.993	27
	INST	2.57	1.089	14
	Total	2.39	1.022	41
ACTm=24-27	DEPT	2.66	.976	59
	INST	2.85	.935	68
	Total	2.76	.955	127
ACTm=28-32	DEPT	2.84	.821	64
	INST	3.09	.900	89
	Total	2.99	.873	153
ACTm=33-36	DEPT	3.75	.500	4
	INST	3.50	.905	12
	Total	3.56	.814	16
Total	DEPT	2.68	.951	161
	INST	2.96	.953	189
	Total	2.83	.961	350

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Calculus III

F	df1	df2	Sig.
1.055	9	340	.396

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+ACT_GRP+DEPTINST+ACT_GRP * DEPTINST

Statistics for Calculus III Using Grades A through D.

Tests of Between-Subjects Effects

Dependent Variable: Calculus III

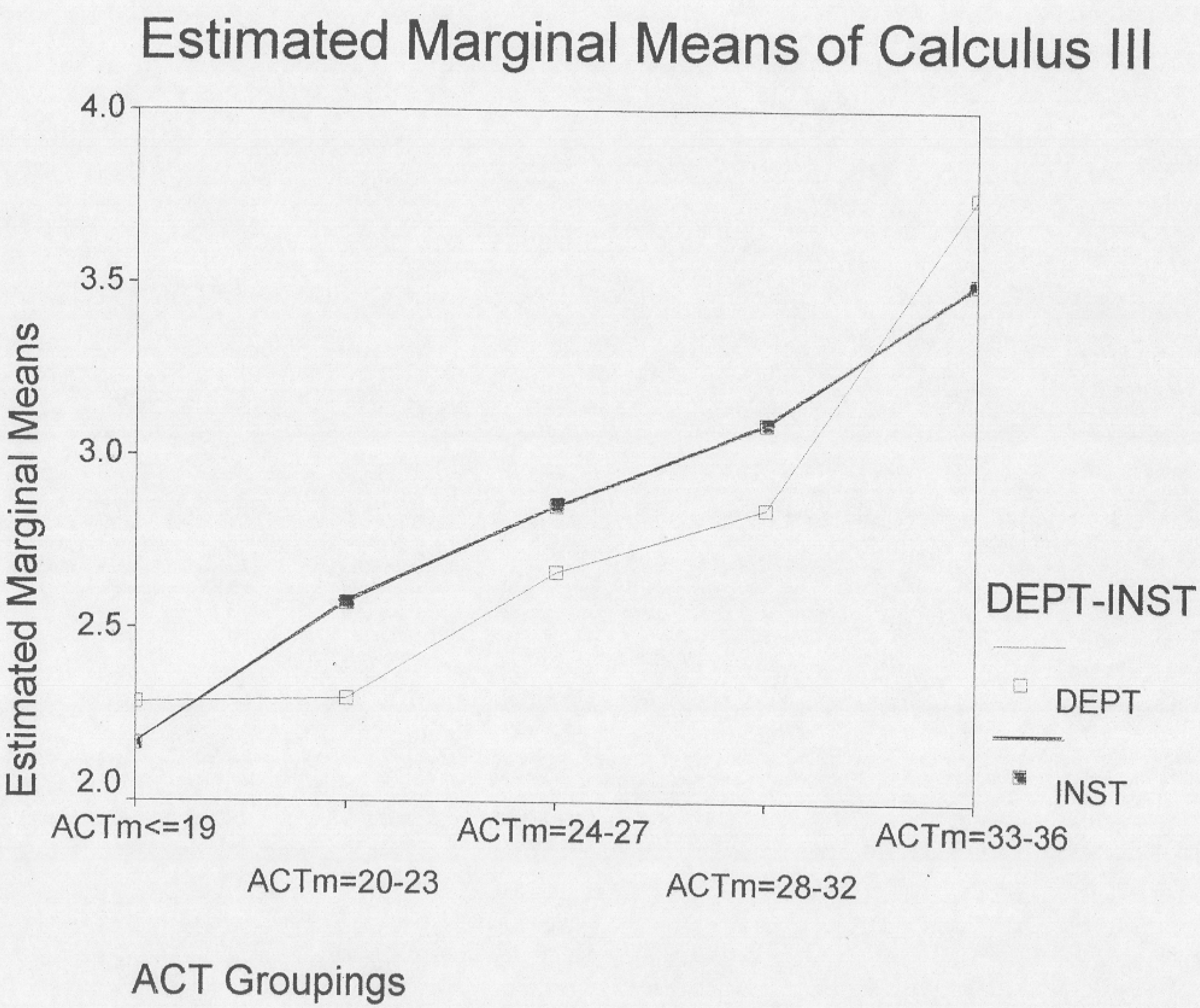
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	29.850[b]	9	3.317	3.855	.000	34.693	.994
Intercept	969.318	1	969.318	1126.582	.000	1126.582	1.000
ACT_GRP	21.536	4	5.384	6.257	.000	25.029	.988
DEPTINST	.146	1	.146	.170	.680	.170	.070
ACT_GRP * DEPTINST	1.069	4	.267	.311	.871	1.243	.120
Error	292.538	340	.860				
Total	3134.000	350					
Corrected Total	322.389	349					

a. Computed using alpha = .05

b. R Squared = .093 (Adjusted R Squared = .069)

Profile Plots

Statistics for Calculus III Using Grades A through D.



Statistics for Calculus IV Using Grades A through D.

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
ACT Groupings	1	ACTm<=19	7
	2	ACTm=20-23	22
	3	ACTm=24-27	95
	4	ACTm=28-32	124
	5	ACTm=33-36	14
DEPT-INST	0	DEPT	117
	1	INST	145

Statistics for Calculus IV Using Grades A through D.

Descriptive Statistics

Dependent Variable: Calculus IV

ACT Groupings	DEPT-INST	Mean	Std. Deviation	N
ACTm<=19	DEPT	2.67	.577	3
	INST	2.25	1.500	4
	Total	2.43	1.134	7
ACTm=20-23	DEPT	2.80	1.082	15
	INST	2.57	1.134	7
	Total	2.73	1.077	22
ACTm=24-27	DEPT	3.10	.889	41
	INST	2.76	.950	54
	Total	2.91	.935	95
ACTm=28-32	DEPT	3.41	.687	54
	INST	3.23	.904	70
	Total	3.31	.818	124
ACTm=33-36	DEPT	4.00	.000	4
	INST	3.20	.789	10
	Total	3.43	.756	14
Total	DEPT	3.22	.842	117
	INST	2.99	.968	145
	Total	3.10	.919	262

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Calculus IV

F	df1	df2	Sig.
2.698	9	252	.005

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+ACT_GRP+DEPTINST+ACT_GRP * DEPTINST

Statistics for Calculus IV Using Grades A through D.

Tests of Between-Subjects Effects

Dependent Variable: Calculus IV

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	22.624[b]	9	2.514	3.199	.001	28.795	.979
Intercept	737.630	1	737.630	938.845	.000	938.845	1.000
ACT_GRP	17.433	4	4.358	5.547	.000	22.189	.976
DEPTINST	3.160	1	3.160	4.022	.046	4.022	.515
ACT_GRP * DEPTINST	1.207	4	.302	.384	.820	1.537	.138
Error	197.991	252	.786				
Total	2731.000	262					
Corrected Total	220.615	261					

a. Computed using alpha = .05

b. R Squared = .103 (Adjusted R Squared = .070)

Profile Plots

Statistics for Calculus IV Using Grades A through D.